

WO 03/104272

PCT/EP03/06096

1/45

Fig. 1 – Sequence information for C-LytA.

SEQ ID NO:1 – amino acid sequence of C-LytA repeat 1

GWQKNDTGYWYVHSD 15

SEQ ID NO:2 – amino acid sequence of C-LytA repeat 2

GSYPKDKFEKINGTWYFDSS 21

SEQ ID NO:3 – amino acid sequence of C-LytA repeat 3

GYMLADRWRKHTDGNWYFDNS 22

SEQ ID NO:4 – amino acid sequence of C-LytA repeat 4

GEMATGWKKIADKWYFNEE 20

SEQ ID NO:5 – amino acid sequence of C-LytA repeat 5

GAMKTGWVKYKDTWYLDKE 21

SEQ ID NO:6 – amino acid sequence of C-LytA repeat 6

GAMVSNAFIQSADGTWYLLKPD 23

SEQ ID NO:7 – amino acid sequence of C-LytA cholin-binding domain

GWQKNDTGYW YVHSDGSYPK DKFEKINGTW YFDSSGYML ADRWRKHTDG NWYWFDNSGE 60

MATGWKKIAD KWYFNEEGA MKTGWVKYKD TWYLDKEG AMVSNAFIQS ADGTGWYLLK 120

PDGTLADRPE FTVEPDGLIT VK 142

SEQ ID NO:8 – amino acid sequence of C-LytA domain from truncated repeat 1 to repeat 6 (as part of our constructs shown in figure 2)

YVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKT

GWVKYKDTWYLDKEGAMVSNAFIQSADGTWYLLKPD

SEQ ID NO:9 – DNA sequence encoding the amino acid sequence of SEQ ID NO:1

ggctggcaga agaatgacac tggtactgg tacgtacatt cagac

SEQ ID NO:10 – DNA sequence encoding the amino acid sequence of SEQ ID NO:2

ggctcttatc caaaagacaa gtttgagaaa atcaatggca cttggtacta ctttgacagt tca

WO 03/104272

PCT/EP03/06096

2/45

SEQ ID NO:11 – DNA sequence encoding the amino acid sequence of SEQ ID NO:3

ggctatatgc ttgcagaccg ctggaggaag cacacagacg gcaactggta ctggttcgac aactca

SEQ ID NO:12 – DNA sequence encoding the amino acid sequence of SEQ ID NO:4

ggcgaaatgg ctacaggctg gaagaaaatc gctgataagt ggtactatct caacgaagaa

SEQ ID NO:13 – DNA sequence encoding the amino acid sequence of SEQ ID NO:5

Ggtgccatga agacaggctg ggtcaagtac aaggacactt ggtactactt agacgctaaa gaa

SEQ ID NO:14 – DNA sequence encoding the amino acid sequence of SEQ ID NO:6

Ggcgccatgg tatcaaatgc ctttatccag tcagcggacg gaacaggctg gtactacctc
aaaccagac

SEQ ID NO:15 – DNA sequence encoding the amino acid sequence of SEQ ID NO:7

ggctggcaga agaatgacac tggctactgg tacgtacatt cagacggctc ttatccaaaa 60
gacaagtttg agaaaatcaa tggcacttgg tactactttg acagttcagg ctatatgctt 120
gcagaccgct ggaggaagca cacagacggc aactgggtact ggttcgacaa ctcaggcgaa 180
atggctacag gctggaagaa aatcgctgat aagtgggtact atttcaacga agaaggtgcc 240
atgaagacag gctgggtcaa gtacaaggac acttgggtact acttagacgc taaagaaggc 300
gccatggtat caaatgcctt tatccagtca gcggacggaa caggctggta ctacctcaa 360
ccagacggaa cactggcaga caggccagaa ttcacagtag agccagatgg cttgattaca 420
gtaaaataa 429

SEQ ID NO:16 – DNA sequence encoding the amino acid sequence of SEQ ID NO:8

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
GTTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTA
TCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGAC

WO 03/104272

PCT/EP03/06096

3/45

FIG. 2. CPC and native Constructs

Construct 1 – coding sequence of CPC-P501₅₁₋₅₅₃ (see plasmid of figure 7 -Y1796)

Protein sequence (SEQ ID NO:27)

MAAA^{R1}YVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWF^{R2}DN^{R3}SGEMATG^{R4}
^{R5}WKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGA^{P2}MOYIKANSKEFIGITEGV^{R6}MVSNAFIQS
^{R5}ADGTGWYYLKPD^{R5}GT^{R5}LADRPEKFMVMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALS
 L GILLSLFLIPRAGWLAGLLCPDRPLELALLILGVLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSV
 YAFMISLGGCLGYLLPAIDWDT^{R5}SALAPYLG^{R5}TQE^{R5}ECFLG^{R5}LLTLIFLTCVAATLLVAEEAALGPTEPAEG
 LSAPSLSPHCCPCRARLAFRNLGALLPRLHQ^{R5}LC^{R5}CRMP^{R5}RTL^{R5}RRLFVAELCSWMALMTFTLFYTD^{R5}FVGE
 GLYQGV^{R5}PRAEP^{R5}TEARRHYDEGVRMGS^{R5}LGLFLQCAISLVFSLVMDRLVQ^{R5}RF^{R5}GTRAVYLASVAAFPV
 AAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSED^{R5}SLMTSF
 LPGPKPGAPFPNGHV^{R5}GAGGSGLLPP^{R5}PALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAF
 LLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSAGGHHHHHH

R1 (plain): aa5-9 (fragment) R4 (bold): aa53-72 P2 (underline): 97-110
 R2 (bold): aa10-30 R5 (plain): aa73-93
 R3 (plain): aa31-52 R6a (bold): aa94-95 R6b (bold): 113-133

Nucleotide sequence (SEQ ID NO:28)

ATGgcggcgcgtTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT
 ACTACTTTGACAGTTCAGGCTATATGCTTGACAGCCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGT
 CGACAACTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGT
 GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAGAAGGCGCCatgcaat
 acatcaaggctaactctaagttcattggtatcactgaaggcgtaTGGTATCAAATGCCTTTATCCAGTCAGC
 GGACGGAACAGGCTGGTACTACTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAaagttcatgtaCatg
 GTGCTGGGCATTGGTCCAGTGCTGGGCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTG
 GACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCC
 AAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCCGATCCCAGGCCCTGGAGCTGGCACTGCTCATCCTGGGC
 GTGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTGTGACCTCTTCCGGG
 ACCCGGACCACTGTGCGCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGCTGCCTGGGCTACCT
 CCTGCCTGCCATTGACTGGGACACAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGAGTGCTCTTTGGC
 CTGCTCACCTCATCTTCCCTACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCA
 CCGAGCCAGCAGAAGGGCTGTGCGCCCCCTCCTGTGCGCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTT

WO 03/104272

PCT/EP03/06096

4/45

CCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGCATGCCCCGCACCTGCGCCGGCTC
TTCGTGGCTGAGCTGTGACAGCTGGATGGCACTCATGACCTTCACGCTGTTTTACACGGATTCGTGGGCGAGG
GGCTGTACCAGGGCGTGTCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGCGTTCCGAT
GGGCAGCCTGGGGCTGTTCTGTCAGTGCGCCATCTCCCTGGTCTTCTCTCTGTCATGGACCGGCTGGTGCAG
CGATTCCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGGCTGCCGGTGCCACATGCCTGT
CCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCC
CTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCTGCCCAAATACCGAGGGGACACTGGAGGT
GCTAGCAGTGAAGACAGCCTGATGACCAGCTTCTGTCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGAC
ACGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCCTGTGAtGTC
CGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGTGGTTCCGGGCCGGGGCATCTGCCTGGACCTCGCC
ATCCTGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCC
AGTCTGTCACTGCCTATATGGTGTCTGCCCGAGGCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGT
ATTTGACAAGAGCGACTTGGCCAAATACTCAGCGggtggacaccatcaccatcaccattaa

Construct 2 – Coding sequence of P501₅₅₋₅₅₃ HIS (control) (yeast strain SC333)

Protein sequence (SEQ ID NO:29)

MVLGIGPVLG	LVCVPLLGSA	SDHWRGRYGR	RRPFIWALSL	GILLSLFLIP	RAGWLAGLLC	60
PDPRLPLELAL	LILGVLLDF	CGQVCFTPLE	ALLSDLFRDP	DHCRQAYSVY	AFMISLGGCL	120
GYLLPAIDWD	TSALAPYLGT	QEECLFGLLT	LIFLTCVAAT	LLVAEEAALG	PTEPAEGLSA	180
PSLSPHCCPC	RARLAFRNLG	ALLPRLHQLC	CRMPRTLRLR	FVAELCSWMA	LMTFTLFYTD	240
FVGEGLYQGV	PRAEPGTEAR	RHYDEGVRMG	SLGLFLQCAI	SLVFSLVMDR	LVQRFGTRAV	300
YLASVAAPV	AAGATCLSHS	VAVVTASAAL	TGFTFSALQI	LPYTLASLYH	REKQVFLPKY	360
RGDTGGASSE	DSLMTSFLPG	PKPGAPFPNG	HVGAGGSGLL	PPPPALCGAS	ACDVSVRVVV	420
GEPTEARVVP	GRGICLDLAI	LDSAFLLSQV	APSLFMGSIV	QLSQSVTAYM	VSAAGLGLVA	480
IYFATQVVFD	KSDLAKYSAG	GHHHHHH				507

Nucleotide sequence (SEQ ID NO:30)

atgTGCTGG	GCATTGGTCC	AGTGTGGGC	CTGGTCTGTG	TCCCGCTCCT	AGGCTCAGCC	60
AGTGACCACT	GGCGTGGACG	CTATGGCCGC	CGCCGGCCCT	TCATCTGGGC	ACTGTCCTTG	120
GGCATCCTGC	TGAGCCTCTT	TCTCATCCCA	AGGGCCGGCT	GGCTAGCAGG	GCTGCTGTGC	180
CCGGATCCCA	GGCCCCCTGA	GCTGGCACTG	CTCATCCTGG	GCGTGGGGCT	GCTGGACTTC	240
TGTGGCCAGG	TGTGCTTCAC	TCCACTGGAG	GCCCTGCTCT	CTGACCTCTT	CCGGGACCCG	300
GACCACTGTC	GCCAGGCCTA	CTCTGTCTAT	GCCTTCATGA	TCAGTCTTGG	GGGCTGCCTG	360
GGCTACCTCC	TGCCTGCCAT	TGACTGGGAC	ACCAGTGCCC	TGGCCCCCTA	CCTGGGCACC	420
CAGGAGGAGT	GCCTCTTTGG	CCTGCTCACC	CTCATCTTCC	TCACCTGCGT	AGCAGCCACA	480
CTGCTGGTGG	CTGAGGAGGC	AGCGTGGGC	CCCACCGAGC	CAGCAGAAGG	GCTGTCGGCC	540
CCCTCCTTGT	CGCCCCACTG	CTGTCCATGC	CGGGCCCGCT	TGGCTTTCCG	GAACCTGGGC	600

WO 03/104272

PCT/EP03/06096

5/45

```

GCCCTGCTTC CCCGGCTGCA CCAGCTGTGC TGCCGCATGC CCCGCACCCT GCGCCGGCTC 660
TTCGTGGCTG AGCTGTGCAG CTGGATGGCA CTCATGACCT TCACGCTGTT TTACACGGAT 720
TTCGTGGGCG AGGGGCTGTA CCAGGGCGTG CCCAGAGCTG AGCCGGGCAC CGAGGCCCGG 780
AGACACTATG ATGAAGGCGT TCGGATGGGC AGCCTGGGGC TGTTCTTGCA GTGCGCCATC 840
TCCCTGGTCT TCTCTCTGGT CATGGACCGG CTGGTGACGC GATTTCGGAC TCGAGCAGTC 900
TATTTGGCCA GTGTGGCAGC TTTCCCTGTG GCTGCCGGTG CCACATGCCT GTCCACAGT 960
GTGGCCGTGG TGACAGCTTC AGCCGCCCTC ACCGGGTTC CCTTCTCAGC CCTGCAGATC 1020
CTGCCCTACA CACTGGCCTC CCTCTACCAC CGGGAGAAGC AGGTGTTCTT GCCCAAATAC 1080
CGAGGGGACA CTGGAGGTGC TAGCAGTGAG GACAGCCTGA TGACCAGCTT CCTGCCAGGC 1140
CCTAAGCCTG GAGCTCCCTT CCCTAATGGA CACGTGGGTG CTGGAGGCAG TGGCCTGCTC 1200
CCACCTCCAC CCGCGCTCTG CGGGGCCTCT GCCTGTGATG TCTCCGTACG TGTGGTGGTG 1260
GGTGAGCCCA CCGAGGCCAG GGTGGTTCCG GGCCGGGGCA TCTGCCTGGA CCTCGCCATC 1320
CTGGATAGTG CCTTCTGCT GTCCAGGTG GCCCATCCC TGTTTATGGG CTCCATTGTC 1380
CAGCTCAGCC AGTCTGTCAC TGCTATATG GTGTCTGCCG CAGGCCTGGG TCTGGTCGCC 1440
ATTTACTTTG CTACACAGGT AGTATTTGAC AAGAGCGACT TGGCCAAATA CTCAGCGgt 1500
ggacaccatc accatcacca ttaa 1524

```

Construct 3 - Coding sequence of natssP501₁₋₃₄ P501₅₁₋₅₅₃ HIS (yeast strain Y1800)

Protein sequence (SEQ ID NO:31)

```

                                R1      R2
MAAVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAAYVHSDGSYPKDKFEKINGTW
                                R3      R4      R5
YYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYYFNEEGAMKTGWVK
                                P2      R6
YKDTWYYLDAKEGAMMOYIKANSKFIGITEGVMVSNAFIQSADGTGWYYLKPDGTLADRPEKFMY
MVLGIGPVLGLVCVPLLGSASDHWGRYGRRRPFIWALSGLLSLFLIPRAGWLAGLLCPDPRPLEL
ALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRAYSVYAFMISLGGCLGYLLPAIDWDTSALAP
YLGTQEELFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR
LHQLCCRMPTLRRLFVAELCSWMALMTFTLFYTDVFGEGLYQGVPRAEPTARRHYDEGVRMG
SLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSA
LQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPA
LCGASACDVSVRVVVGEPTARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVVS
AAGLGLVAIFYFATQVVFDKSDLAKYSAGGHHHHHH

```

R1 (plain): aa38-42 (fragment)

R4 (bold): aa77-106

P2 (underline): 130-143

R2 (bold): aa43-64

R5 (plain): aa107-126

R3 (plain): aa65-76

R6a (bold): aa127-128

R6b (bold): aa146-166

WO 03/104272

PCT/EP03/06096

6/45

natss stands for native signal sequence

Nucleotide sequence (SEQ ID NO:32)

ATGgcGGCCGTGCAGAGGCTATGGGTATCGAGACTGCTAAGACACCGCAAAGCTCAGTTGTTGTTGGTTAACT
 TGTTGACCTTCGGGCTGGAAGTCTGTTTGGCggcgctTACGTACATTCCGACGGCTCTTATCCAAAAGACAA
 GTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTTCAGGCTATATGCTTGCAGACCGCTGGAGGAAG
 CACACAGACGGCAACTGGTACTGGTTCGACAACCTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATA
 AGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTT
 AGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactctaagttcattggtatcactgaaggcgctcATG
 GTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACTCTCAAACCAGACGGAACACTGGCAG
 ACAGGCCAGAAaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTCTGGGCCCTGGTCTGTGTCCCGCTCCT
 AGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACTGTCCTTGGGC
 ATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCAGGCCCC
 TGGAGCTGGCACTGCTCATCTGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGA
 GGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTGCCAGGCTACTCTGTCTATGCCTTCATGATC
 AGTCTTGGGGGCTGCCTGGGCTACCTCCTGCGTGCATTTGACTGGGACACCAGTGGCCTGGCCCCCTACCTGG
 GCACCCAGGAGGAGTGCTCTTTGGCCTGCTCACCCCTCATCTTCTCACCTGCGTAGCAGCCACACTGCTGGT
 GGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTGCGCCCCCTCCTTGTCGCCCCACTGC
 TGTCCATGCCGGGCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCC
 GCATGCCCCGCACCCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCT
 GTTTTACACGGATTTCTGTGGGCGAGGGGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCCGG
 AGACACTATGATGAAGGCGTTCCGATGGGCAGCCTGGGGCTGTTCTTGCAGTGCGCCATCTCCCTGGTCTTCT
 CTCTGGTTCATGGACCGGCTGGTGCAGCGATTCCGCACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTTCCC
 TGTGGCTGCCGGTGCCACATGCCTGTCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTC
 ACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGAGAGAAGCAGGTGTTCTTGC
 CCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCTTGCAGGCCCTAA
 GCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCGCGCTC
 TGCGGGGCCCTCTGCCTGTGAtGTCTCCGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGTGGTTCGGG
 GCCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCCATCCCTGTT
 TATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGAGGCCTGGGTCTGGTC
 GCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGgtggacaccatc
 accatcaccattaa

Construct 4 - Coding sequence of alphapreCPC-P501₅₁₋₅₅₃ HIS (yeast strain Y1802)

Protein sequence (SEQ ID NO:33)

Alpha-pre	signal	R1	R2	R3
MAARFPSIFTAVLFAASSALAAA				
		YVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWKHTDGNWYWF		
R4		R5	E2	
NSGEMATGWKKIADKWYFNEEGAMKTGWVKYKDTWYYLDAKEGAM		MOYIKANSKFIGITEGVMVSNAFI		

WO 03/104272

PCT/EP03/06096

7/45

R6

QSADGTGWYLLKPD**Q**TLADRPEKFMVMVLGIGPVLGLVCPVLLGSASDHWRGRYGRRRRPFIWALSLGILLSLF
 LI PRAGWLAGLLCPDRPLELALLILGVGLLDFCGQVCFPLEALLSDLFRDPDHCRAQSVYAFMISLGGCL
 GYLLPAIDWDTSALAPYLGTQEELFGLLTLIFLTCAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRAR
 LAFRNLGALLPRLHQLCCRMPTLRRLFVAELCSWMALMTFTLFYTD FVGEGLYQGVPRAPGTEARRHYDEG
 VRMGSGLFLQCAISLVFSLVMDRLVQRFGRVAVLASVAAPVVAAGATCLSHSVAVVTASAALTGFTFSALQ
 ILPYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSFLPGPKGAPFPNGHVGAGGSGLLPPPPALCGASAC
 DVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFAT
 QVVFDKSDLAKYSAGGHHHHH

Alpha-pre signal (bold): aa4-22

R1 (plain): aa24-28 (fragment)

R4 (bold): aa72-91

P2 (underline): 116-129

R2 (bold): aa29-49

R5 (plain): aa92-112

R3 (plain): aa50-71

R6a (bold): aa113-114

R6b (bold): aa132-152

Alphapre stands for alpha pre signal sequence

Nucleotide sequence (SEQ ID NO:34)

TACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACA
 GTTCAGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGG
 CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACA
 GGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaatacatcaaggcta
actctaagttcattggtatcactgaaggcgctATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGG
 CTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAA

ATGgcGGCCAGATTTCTTCAATTTTACTGCACTTTTATTCGCAGCATCCTCCGCATTAGCggccgctTACG
 TACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTC
 AGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAA
 ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCT
 GGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCatgcaatacatcaaggctaactc
taagttcattggtatcactgaaggcgctATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGG
 TACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAgctggtattacttaegttccaccattgttgt
 tggaagttggtgtgaagaaaagttcatgtaCatgGTGCTGGGCATTGGTCCAGTGCTGGGCTGGTCTGTGT
 CCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGACGCTATGGCCGCCGCCGCCCTTCATCTGGGCACTG
 TCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATC
 CCAGGCCCTTGAGCTGGCACTGCTCATCTGGCGTGCGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCAC
 TCCACTGGAGGCCCTGCTCTGACCTCTTCCGGGACCCGACCCTGTGCGCCAGGCCTACTCTGTCTATGCT
 TCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAGTGCCCTGGCCCC
 CTACCTGGGCACCCAGGAGGAGTGCTCTTTGGCTGCTCACCTCATCTTCTCACCTGCGTAGCAGCCACA

WO 03/104272

PCT/EP03/06096

8/45

CTGCTGGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGCCCCCTCCTTGTTCGC
 CCCACTGCTGTCCATGCGGGCCCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCACT
 GTGCTGCCGATGCCCCGACCCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACC
 TTCACGCTGTTTTACACGATTTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCG
 AGGCCCCGAGACACTATGATGAAGGCGTTCGGATGGGCGAGCCTGGGGCTGTTCTTGCACTGCGCCATCTCCCT
 GGTCTTCTCTCTGCTCATGGACCGGCTGGTGCAGCGATTTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCA
 GCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCA
 CCGGGTTACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACGGGAGAAGCAGGT
 GTTCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCTTGCCA
 GGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGAGTGGCCTGCTCCACCTCCAC
 CCGCGCTCTGCGGGGCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGT
 GGTTCGGGCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCCA
 TCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGACGGCCTGG
 GTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGgttg
 acaccatcaccatcaccattaa

Construct 5 - Coding sequence of alphaprepro-P501₅₁₋₅₅₃ HIS (in plasmid pRIT 15068 and yeast strain Y1790)

Protein sequence (SEQ ID NO:35)

MSFLNFTAVL	FAASSALAAP	VNTTTEDETA	QIPAEAVIGY	SDLEGDFDVA	VLPFSNSTNN	60
GLLFINTTIA	SIAAKEEGVS	LEKREAEAMV	LGIGPVLGLV	CVPLLGSASD	HWRGRYGRRR	120
PFIWALSIGI	LLSLFLIPRA	GWLAGLLCPD	PRPLELALLI	LGVGLLDPCG	QVCFTPLEAL	180
LSDLFRDPDH	CRQAYSVYAF	MISLGGCLGY	LLPAIDWDTS	ALAPYLGTQE	ECLFGLLTLI	240
FLTCAATLL	VAEEAALGPT	EPAEGLSAPS	LSPHCCPCRA	RLAFRNLGAL	LPRLHQLCCR	300
MPRTLRLRFV	AELCSWMALM	TFTLFYTDV	GEGLYQGVPR	AEPGTEARRH	YDEGVRMGSL	360
GLFLQCAISL	VFSLVMDRLV	QRFGTRAVYL	ASVAAFPVAA	GATCLSHSVA	VVTASAALTG	420
FTFSALQILP	YTLASLYHRE	KQVFLPKYRG	DTGGASSEDS	LMTSFLPGPK	PGAPFPNGHV	480
GAGGSGLLPP	PPALCGASAC	DVSVRVVVE	PTEARVVPGR	GICLDLAILD	SAFLLSQVAP	540
SLFMGSIVQL	SQSVTAYMVS	AAGLGLVAIY	FATQVVFDKS	DLAKYSAGGH	HHHHH	595

Nucleotide sequence (SEQ ID NO:36)

ATGAGTTTCC	TCAATTTTAC	TGCAGTTTTA	TTCGCAGCAT	CCTCCGCATT	AGCTGCTCCA	60
GTCAACACTA	CAACAGAAGA	TGAAACGGCA	CAAATTCGG	CTGAAGCTGT	CATCGGTTAC	120
TCAGATTAG	AAGGGGATT	CGATGTGCT	GTTTGTCCAT	TTTCCAACAG	CACAAATAAC	180
GGGTTATTGT	TTATAAATAC	TACTATTGCC	AGCATTGCTG	CTAAAGAAGA	AGGGGTATCT	240
CTCGAGAAAA	GAGAGGCTGA	AGCCatgGTG	CTGGGCATTG	GTCCAGTGCT	GGGCCTGGTC	300
TGTGTCCCGC	TCCTAGGCTC	AGCCAGTGAC	CACTGGCGTG	GACGCTATGG	CCGCCGCCGG	360

WO 03/104272

PCT/EP03/06096

9/45

```

CCCTTCATCT GGGCACTGTC CTTGGGCATC CTGCTGAGCC TCTTTCTCAT CCCAAGGGCC 420
GGCTGGCTAG CAGGGCTGCT GTGCCCCGAT CCCAGGCCCC TGGAGCTGGC ACTGCTCATC 480
CTGGGCGTGG GGCTGCTGGA CTTCTGTGGC CAGGTGTGCT TCACTCCACT GGAGGCCCTG 540
CTCTCTGACC TCTTCCGGGA CCCGGACCAC TGTCGCCAGG CCTACTCTGT CTATGCCTTC 600
ATGATCAGTC TTGGGGGCTG CCTGGGCTAC CTCCTGCCTG CCATTGACTG GGACACCAGT 660
GCCCTGGCCC CCTACCTGGG CACCCAGGAG GAGTGCTCTT TTGGCCTGCT CACCCTCATC 720
TTCCTCACCT GCGTAGCAGC CACACTGCTG GTGGCTGAGG AGGCAGCGCT GGGCCCCACC 780
GAGCCAGCAG AAGGGCTGTC GGCCCCCTCC TTGTCGCCCC ACTGCTGTCC ATGCCGGGCC 840
CGCTTGCTT TCCGGAACCT GGGCGCCCTG CTTCCCCGGC TGCACCAGCT GTGCTGCCGC 900
ATGCCCCGCA CCCTGCGCCG GCTCTTCGTG GCTGAGCTGT GCAGCTGGAT GGCACATCAG 960
ACCTTCACGC TGTTTACAC GGATTTCGTG GCGGAGGGGC TGTACCAGGG CGTGCCGAGA 1020
GCTGAGCCGG GCACCGAGGC CCGGAGACAC TATGATGAAG GCGTTCGGAT GGGCAGCCTG 1080
GGGCTGTTCC TGCACTGCGC CATCTCCCTG GTCTTCTCTC TGGTCATGGA CCGGCTGGTG 1140
CAGCGATTCT GCACTCGAGC AGTCTATTG GCCAGTGTGG CAGCTTTCCC TGTGGCTGCC 1200
GGTGCCACAT GCCTGTCCCA CAGTGTGGCC GTGGTGACAG CTTAGCCGC CCTCACCAGG 1260
TTCACCTTCT CAGCCCTGCA GATCCTGCCC TACACACTGG CCTCCCTCTA CCACCGGGAG 1320
AAGCAGGTGT TCCTGCCCAA ATACCGAGGG GACACTGGAG GTGCTAGCAG TGAGGACAGC 1380
CTGATGACCA GCTTCCTGCC AGGCCCTAAG CCTGGAGCTC CTTCCCTAA TGGACACGTG 1440
GGTGTGGAG GCAGTGGCCT GCTCCACCT CCACCCGCGC TCTGCGGGGC CTCTGCCTGT 1500
GATGTCTCCG TACGTGTGGT GGTGGGTGAG CCCACCGAGG CCAGGGTGGT TCCGGGCCGG 1560
GGCATCTGCC TGGACCTCGC CATCCTGGAT AGTGCCCTCC TGCTGTCCCA GGTGGCCCCA 1620
TCCCTGTTTA TGGGCTCCAT TGTCCAGCTC AGCCAGTCTG TCACTGCCTA TATGGTGTCT 1680
GCCGAGGCC TGGGTCTGGT CGCCATTTAC TTTGCTACAC AGGTAGTATT TGACAAGAGC 1740
GACTTGGCCA AATACTCAGC Gggtggacac catcaccatc accattaa 1788

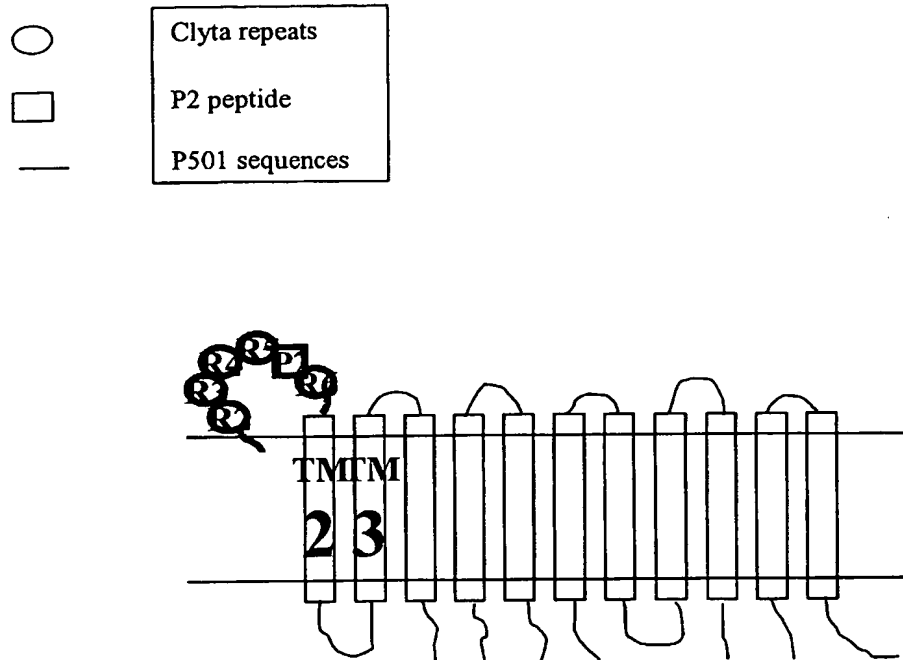
```

WO 03/104272

PCT/EP03/06096

10/45

FIG. 3. Structure of CPC-p501 His fusion protein expressed in *S. cerevisiae*



WO 03/104272

PCT/EP03/06096

11/45

FIG. 4. Primary structure of CPC-P501 His fusion protein (SEQ ID NO.41)

```

MAAAYVHSDG SYPKDKFEKI NGTWYYFDSS GYMLADRWRK HTDGNWYWFD NSGEMATGWK 60
KIADKWYYFN EEGAMKTGWV KYKDTWYYLD AKEGAMQYIK ANSKFIGITE GVMVSNAFIQ 120
SADGTGWYYL KPDGTLADRP EKFMVMVLGI GPVLGLVCVP LLGSASDHWR GRYGRRRPFI 180
WALSLGILLS LFLIPRAGWL AGLLCPDPRP LELALLILGV GLLDFCGQVC FTPLEALLSD 240
LFRDPDHCRQ AYSVYAFMIS LGGCLGYLLP AIDWDTSALA PYLGTQEECL FGLLTLIFLT 300
CVAATLLVAE EAALGPTEPA EGLSAPSLSP HCCPCRARLA FRNLGALLPR LHQLCCRMPR 360
TLRRLFVAEL CSWMALMTFT LFYTDFVGEG LYQGVPRAP GTEARRHYDE GVRMGSLGLF 420
LQCAISLVFS LVMDRLVQRF GTRAVYLASV AAFPVAAGAT CLSHSVAVVT ASAALTGFTF 480
SALQILPYTL ASLYHREKQV FLPKYRGDTG GASSEDSLMT SFLPGPKPGA PFPNGHVGAG 540
GSGLLPPPPA LCGASACDVS VRVVVGEPTE ARVVPGRGIC LDLAILDSAF LLSQVAPSLF 600
MGSIVQLSQS VTAYMVSAAG LGLVAIYFAT QVVFDKSDLA KYSAGGHHHH HH 652

```

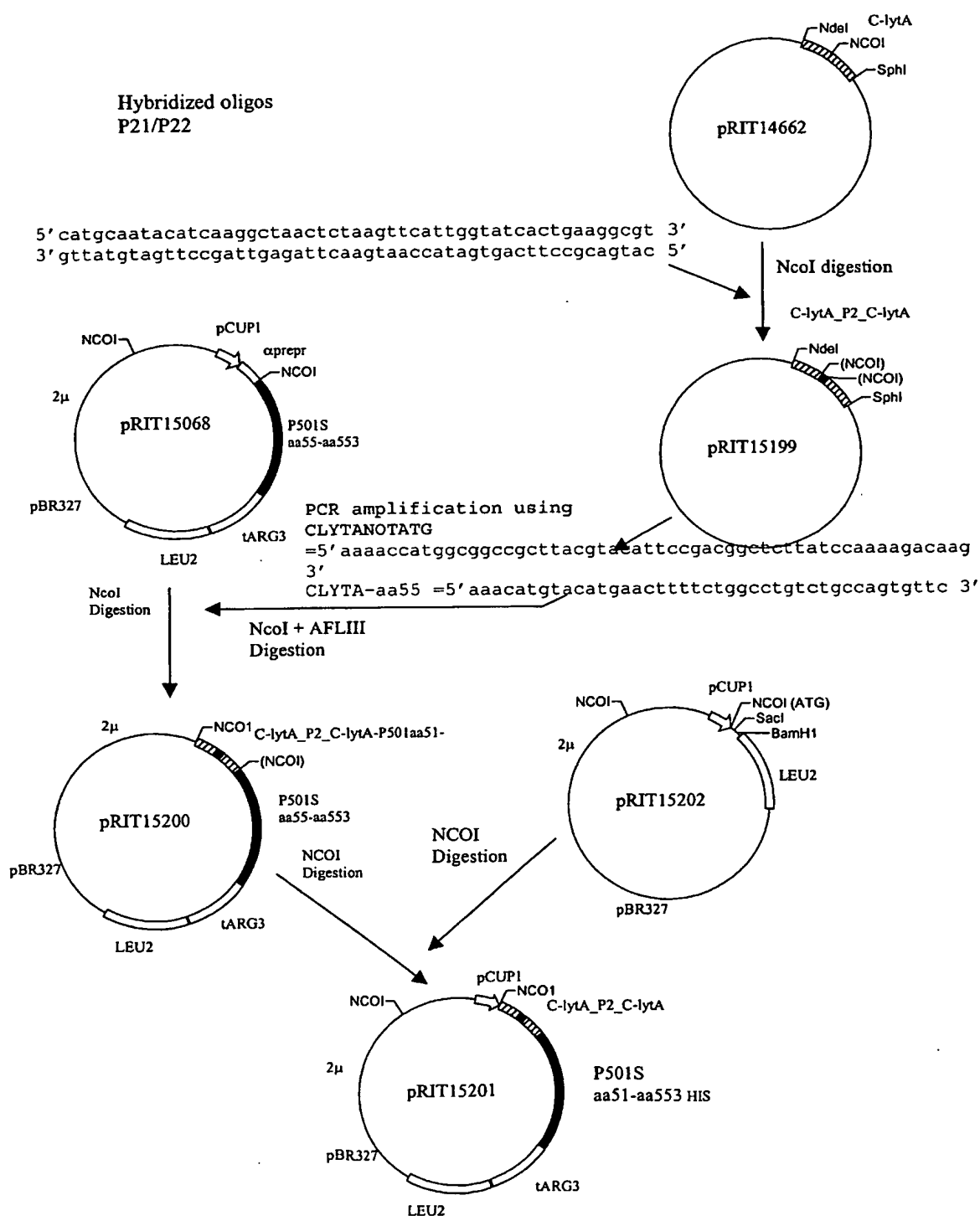
WO 03/104272

PCT/EP03/06096

12/45

FIG. 5. Nucleotide sequence of CPC P501 His(pRIT15201) (SEQ ID NO.42)

```
ATGGCGGCCG CTTACGTACA TTCCGACGGC TCTTATCCAA AAGACAAGTT TGAGAAAATC 60
AATGGCACTT GGTACTACTT TGACAGTTCA GGCTATATGC TTGCAGACCG CTGGAGGAAG 120
CACACAGACG GCAACTGGTA CTGGTTCGAC AACTCAGGCG AAATGGCTAC AGGCTGGAAG 180
AAAATCGCTG ATAAGTGGTA CTATTTCAAC GAAGAAGGTG CCATGAAGAC AGGCTGGGTC 240
AAGTACAAGG ACACTTGGTA CTACTTAGAC GCTAAAGAAG GCGCCATGCA ATACATCAAG 300
GCTAACTCTA AGTTCATTGG TATCACTGAA GCGTCATGG TATCAAATGC CTTTATCCAG 360
TCAGCGGACG GAACAGGCTG GTACTACCTC AAACCAGACG GAACACTGGC AGACAGGCCA 420
GAAAAGTTCA TGTACATGGT GCTGGGCATT GGTCCAGTGC TGGGCCTGGT CTGTGTCCCG 480
CTCCTAGGCT CAGCCAGTGA CCACTGGCGT GGACGCTATG GCCGCCGCCG GCCCTTCATC 540
TGGGCACTGT CCTTGGGCAT CCTGCTGAGC CTCTTTCTCA TCCCAAGGGC CGGCTGGCTA 600
GCAGGGCTGC TGTGCCCGGA TCCCAGGCCG CTGGAGCTGG CACTGCTCAT CCTGGGCGTG 660
GGGCTGTGTG ACTTCTGTGG CCAGGTGTGC TTCACTCCAC TGGAGGCCCT GCTCTCTGAC 720
CTCTTCCGGG ACCCGGACCA CTGTCGCCAG GCCTACTCTG TCTATGCCTT CATGATCAGT 780
CTTGGGGGCT GCCTGGGCTA CCTCCTGCCT GCCATTGACT GGGACACCAG TGCCCTGGCC 840
CCCTACCTGG GCACCCAGGA GGAGTGCCTC TTTGGCCTGC TCACCCTCAT CTTCTCACC 900
TGCGTAGCAG CCACACTGCT GGTGGCTGAG GAGGCAGCGC TGGGCCCCAC CGAGCCAGCA 960
GAAGGGCTGT CGGCCCCCTC CTTGTGCGCC CACTGCTGTC CATGCCGGGC CCGCTTGGCT 1020
TTCCGGAACC TGGGCGCCCT GCTTCCCCGG CTGCACCAGC TGTGCTGCCG CATGCCCGGC 1080
ACCCTGCGCC GGCTCTTCGT GGCTGAGCTG TGCAGCTGGA TGGCACTCAT GACCTTCACG 1140
CTGTTTTACA CGGATTTCGT GGGCGAGGGG CTGTACCAGG GCGTGCCAG AGCTGAGCCG 1200
GGCACCGAGG CCCGGAGACA CTATGATGAA GGCGTTCGGA TGGGCAGCCT GGGGCTGTTC 1260
CTGCAGTGCG CCATCTCCCT GGTCTTCTCT CTGGTCATGG ACCGGCTGGT GCAGCGATTG 1320
GGCACTCGAG CAGTCTATTT GGCCAGTGTG GCAGCTTTCC CTGTGGCTGC CGGTGCCACA 1380
TGCTGTCCG ACAGTGTGGC CGTGGTGACA GCTTCAGCCG CCCTCACCAG GTTCACCTTC 1440
TCAGCCCTGC AGATCCTGCC CTACACACTG GCCTCCCTCT ACCACCGGGA GAAGCAGGTG 1500
TTCTGCCCCA AATACCGAGG GGACACTGGA GGTGCTAGCA GTGAGGACAG CCTGATGACC 1560
AGCTTCTGCG CAGGCCCTAA GCCTGGAGCT CCCTTCCCTA ATGGACACGT GGGTGTCTGA 1620
GGCAGTGGCC TGCTCCACCC TCCACCCGCG CTCTGCGGGG CCTCTGCCTG TGATGTCTCC 1680
GTACGTGTGG TGGTGGGTGA GCCCACCGAG GCCAGGGTGG TTCCGGGCCG GGGCATCTGC 1740
CTGGACCTCG CCATCCTGGA TAGTGCCTTC CTGCTGTCCC AGGTGGCCCC ATCCCTGTTT 1800
ATGGGCTCCA TTGTCCAGCT CAGCCAGTCT GTCACTGCCT ATATGGTGTC TGCCGCAGGC 1860
CTGGGTCTGG TCGCCATTTA CTTTGCTACA CAGGTAGTAT TTGACAAGAG CGACTTGGCC 1920
AAATACTCAG CGGGTGGACA CCATCACCAT CACCATTAA 1959
```

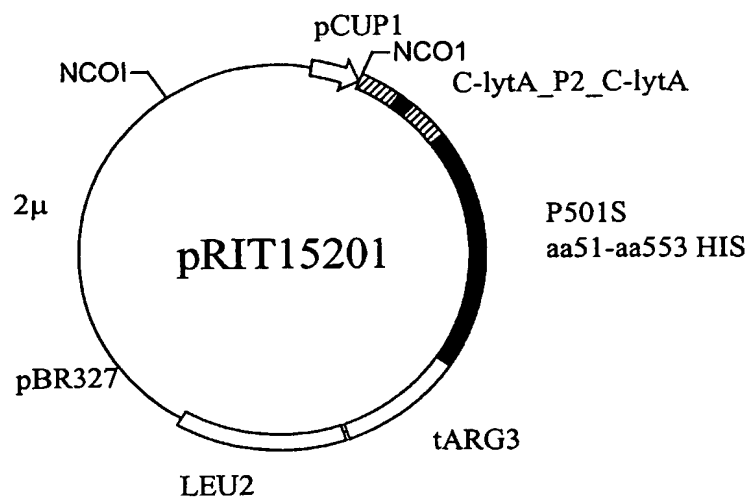


WO 03/104272

PCT/EP03/06096

14/45

FIG. 7. Plasmid map of pRIT15201

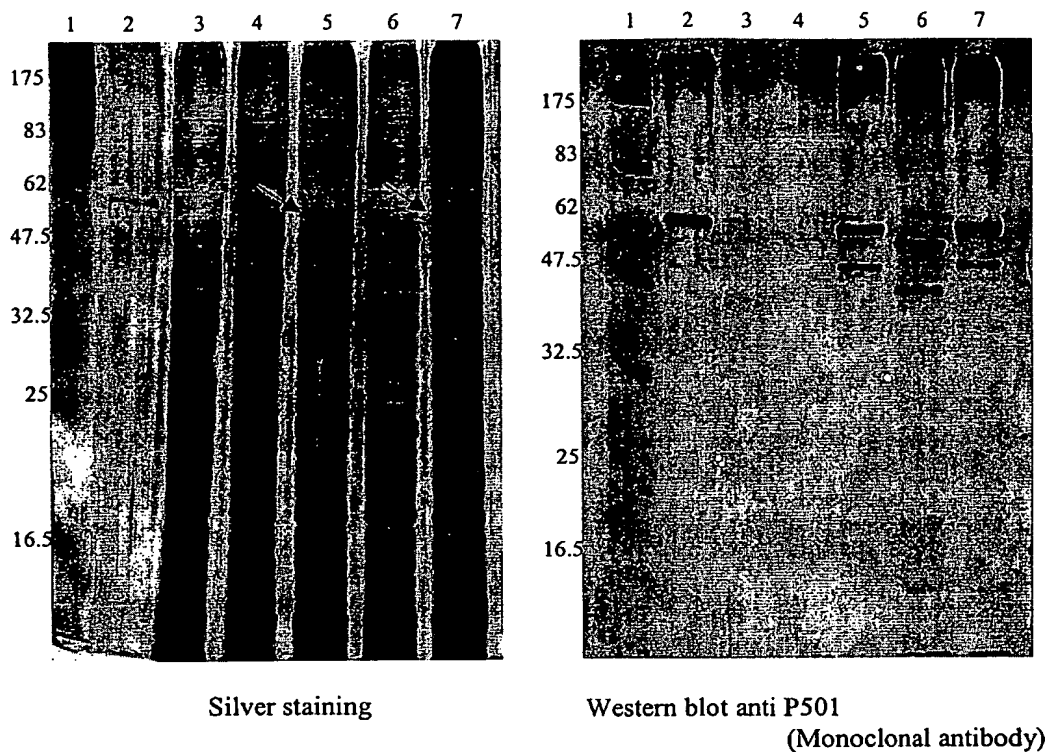


WO 03/104272

PCT/EP03/06096

15/45

FIG. 8. Comparative expression of CPC P501 and P501 in *S.cerevisiae* strain DC5 (gel Laemmli 10%)



- 1 MW Biolabs (175/83/62/47.5/32.5/16.5 Kda)
- 2 Y1796 purified
- 3 Y1795 Crude Extract (negative control)
- 4 SC333 Crude Extract
- 5 Y1796 Crude Extract
- 6 Y1790 Crude Extract
- 7 Y1802 Crude Extract

WO 03/104272

PCT/EP03/06096

16/45

FIG. 9A.

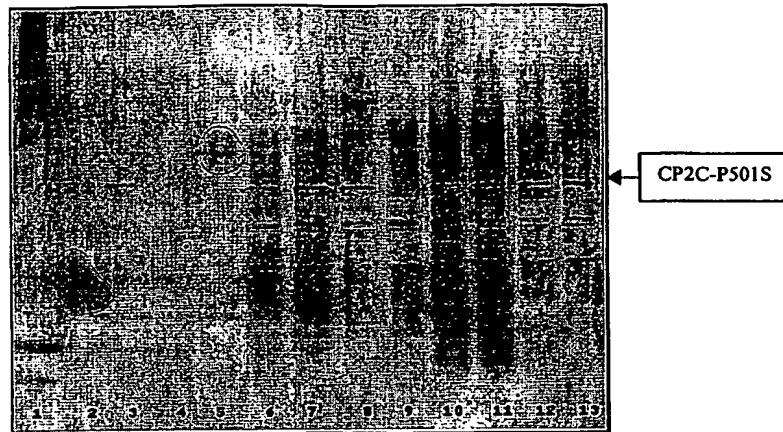
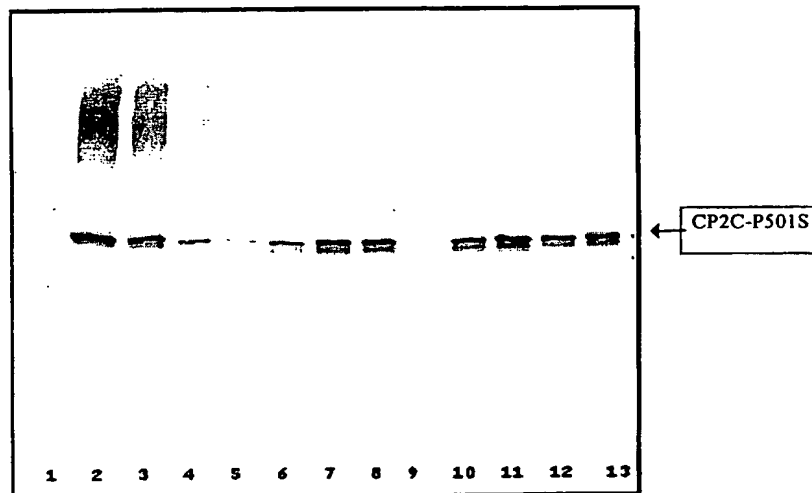


FIG. 9B.



- 1 - Molecular Weight Marker (Biolabs - Grow Range) 175; 83; 62; 47.5; 32.5; 25; 16.5; 6.5 kD - 10
- 2 - Purified Reference CP2CP501S/12 135 ng
- 3 - Purified Reference CP2CP501S/12 67.8 ng
- 4 - Purified Reference CP2CP501S/12 33.9 ng
- 5 - Purified Reference CP2CP501S/12 16.9 ng
- 6 - Fermentation PRO119-21h30
- 7 - Fermentation PRO124-21h30
- 8 - Fermentation PRO124-22h30
- 9 - Fermentation PRO127-0 h
- 10 - Fermentation PRO127-4 h
- 11 - Fermentation PRO127-6 h
- 12 - Fermentation PRO127-22h20
- 13 - Fermentation PRO127-22h45

WO 03/104272

PCT/EP03/06096

17/45

FIG. 10. Purification scheme of CPC-P501-His produced by Y1796.

<i>S. Cerevisiae</i> cells	
↓	
Dyno-mill disruption	OD 120 / 2 passes / 20 mM Tris pH 8.5 - 5 mM EDTA
↓	
Centrifugation	12.000 g / RT / 90 min (supernatant discarded)
↓	
Pellet washing step 1	20 mM Tris pH 8.5 - 0.15 M NaCl - 2.0 M Guanidine.HCl - 0.1% Empigen (30 min / RT)
↓	
Centrifugation	12.000 g / RT / 60 min (supernatant discarded)
↓	
Pellet washing step 2	20 mM Tris pH 8.5 - 0.15 M NaCl - 4.0 M Urea
↓	
Centrifugation	12.000 g / RT / 30 min (supernatant discarded)
↓	
Solubilisation / Reduction	20 mM Tris pH 8.5 - 0.15 M NaCl - 8.0 M Urea - 1% SDS - 0.2 M Glutathion (60 min / RT)
↓	
Centrifugation	12.000 g / RT / 30 min (pellet discarded)
↓	
Carbamidomethylation	0.3 M Iodoacetamide (30 min / RT / in the dark) / pH adjusted to 8.5 (with 5 M NaOH solution) before incubation
↓	
R/C Supernatant	
↓	
10-fold dilution and pH adjustment (8.5)	<u>Dilution buffer</u> : 20 mM Tris pH 8.5 - 1 M NaCl - 8.0 M Urea
↓	
Immobilised metal ion affinity chromatography on Ni²⁺-Chelating Sepharose FF (Amersham) (10x25 cm column – 2000 ml)	<u>Equilibration buffer</u> : 20 mM Tris pH 8.5 - 0.9 M NaCl - 8.0 M Urea - 0.1% SDS <u>Washing buffers</u> : 1) Equilibration buffer 2) 20 mM Tris pH 8.5 - 0.15 M NaCl - 8.0 M Urea - 0.1% SDS 3) 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80

WO 03/104272

PCT/EP03/06096

18/45

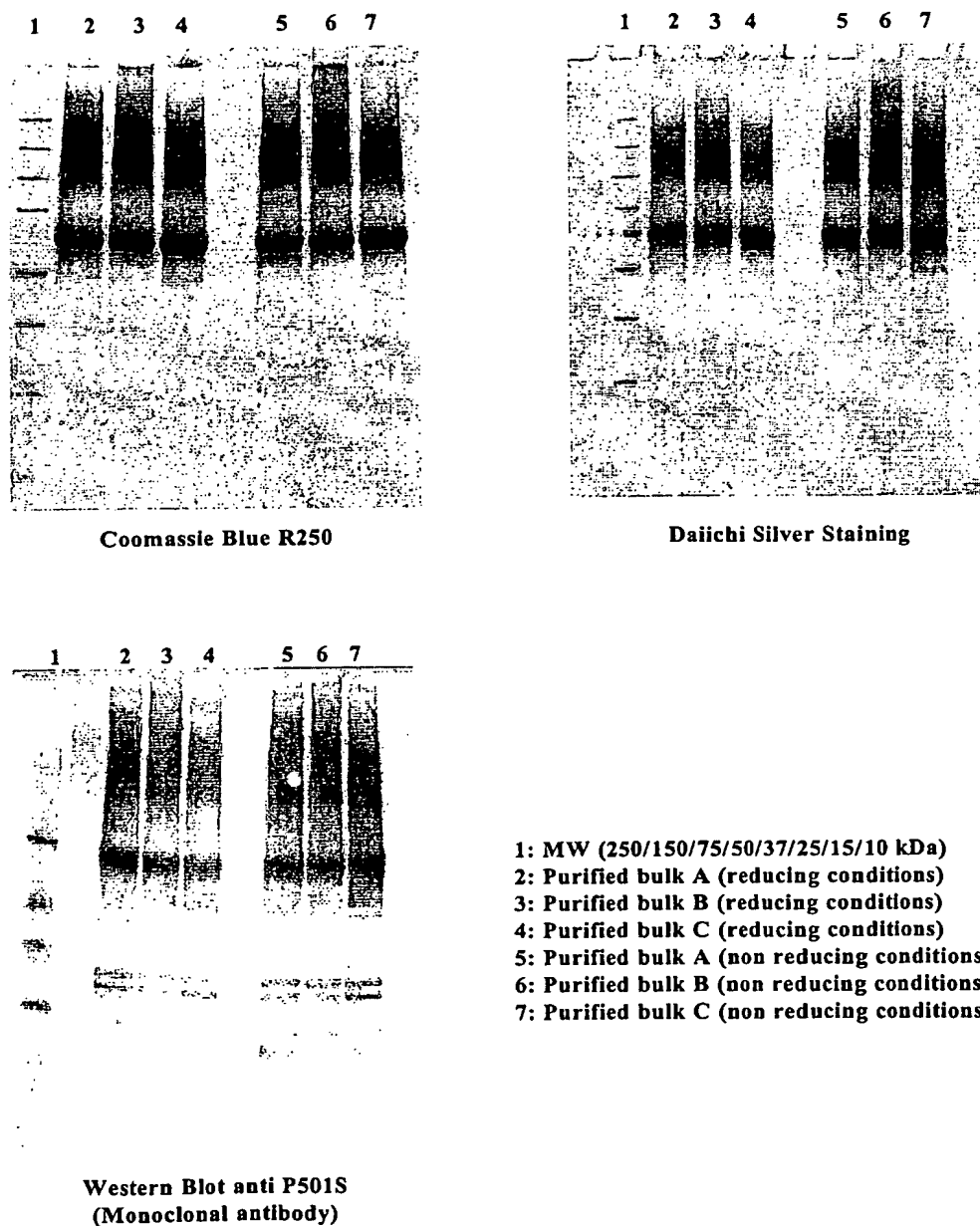
	<u>Elution buffer</u> : 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 - 0.5 M Imidazole
↓	
2-fold dilution and pH adjustment (10.0)	20 mM Piperazine pH 10.0 - 8.0 M Urea - 0.1% Tween 80
↓	
Anion exchange chromatography on Q Sepharose FF (Amersham) (2,6 x 6.5 cm column - 35 ml)	<u>Equilibration buffer</u> : 20 mM Piperazine pH 10.0 - 8.0 M Urea - 0.1% Tween 80 <u>Washing buffers</u> : 1) Equilibration buffer 2) 20 mM Tris pH 8.5 - 8.0 M Urea - 0.1% Tween 80 <u>Elution buffer</u> : 20 mM Tris pH 7.5 - 8.0 M Urea - 0.1% Tween 80 - 0.5 M NaCl
↓	
Concentration/Diafiltration (Pall - Omega 10 kDa - 200 cm²)	+/- 3-fold concentration <u>Diafiltration buffer</u> : Tris 20 mM pH 7.5
↓	
Sterile filtration (Millipore - Millex GV 0.22µm)	
↓	
Purified bulk	<u>Final buffer</u> : 20 mM Tris pH 7.5 - +/- 0.3% Tween 80
↓	
Storage -20°C	

WO 03/104272

PCT/EP03/06096

19/45

FIG. 11. Pattern of CPC P501 His purified protein (4-12% Novex Nu-Page polyacrylamide precasted gels)



WO 03/104272

PCT/EP03/06096

20/45

FIG. 12. Native full-length P501S sequence (SEQ ID NO:17 & 43)

Nucleotide sequence: SEQ ID NO.17

Polypeptide sequence: SEQ ID NO.43

```

#####
GCCACCATGGTCCAGAGGCTGTGGGTGAGCCGCTGCTGCGGCACCGG
      M V Q R L W V S R L L R H R   14

AAAGCCCAGCTCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTTGGCCGCA
K A Q L L L V N L L T F G L E V C L A A   34

GGCATCACCTATGTGCCGCTCTGCTGCTGGAAGTGGGGTAGAGGAGAAGTTCATGACC
G I T Y V P P L L L E V G V E E K F M T   54

ATGGTGTCTGGGCATTGGTCCAGTGTGGGCTGGTCTGTGTCCCCTCCTAGGCTCAGCC
M V L G I G P V L G L V C V P L L G S A   74

AGTGACCACTGGCGTGGACGCTATGGCCGCCCGGCCCTTCATCTGGGCACTGTCCTTG
S D H W R G R Y G R R R P F I W A L S L   94

GGCATCCTGTGAGCCTCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGC
G I L L S L F L I P R A G W L A G L L C   114

CCGGATCCCAGGCCCCCTGGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACTTC
P D P R P L E L A L L I L G V G L L D F   134

TGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGGACCCG
C G Q V C F T P L E A L L S D L F R D P   154

GACCACTGTGCCAGGCCTACTCTGTCTATGCCTTCATGATCAGTCTTGGGGGCTGCCTG
D H C R Q A Y S V Y A F M I S L G G C L   174

GGTACCTCCTGCCTGCCATTGACTGGGACACCACTGCCCTGGCCCCCTACCTGGGCACC
G Y L L P A I D W D T S A L A P Y L G T   194

CAGGAGGAGTGCCTCTTTGGCCTGCTCACCCCTCATCTTCCTCACCTGCGTAGCAGCCACA
Q E E C L F G L L T L I F L T C V A A T   214

CTGCTGGTGGCTGAGGAGGAGCGCTGGGCCCCACCGAGCCAGCAGAAGGGCTGTGCGCC
L L V A E E A A L G P T E P A E G L S A   234

CCCTCCTTGTGCCCCACTGCTGTCCATGCCGGGCGCGCTTGGCTTTCCGGAACCTGGGC
P S L S P H C C P C R A R L A F R N L G   254

GCCCTGCTTCCCCGGCTGCACCAGCTGTGCTGCCGCATGCCCCGCACCCTGCGCCGGCTC
A L L P R L H Q L C C R M P R T L R R L   274

TTCGTGGCTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTACACGGAT
F V A E L C S W M A L M T F T L F Y T D   294

TTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCCGG
F V G E G L Y Q G V P R A E P G T E A R   314

AGACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCTGCACTGCGCCATC
R H Y D E G V R M G S L G L F L Q C A I   334

TCCCTGGTCTTCTCTCTGTCATGGACCGGCTGGTGCAGCGATTGCGCACTCGAGCAGTC
S L V F S L V M D R L V Q R F G T R A V   354

```

WO 03/104272

PCT/EP03/06096

21/45

TATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGT
Y L A S V A A F P V A A G A T C L S H S 374

GTGGCCGTGGTGACAGCTTCAGCCGCCCTCACC GGTTACCTTCTCAGCCCTGCAGATC
V A V V T A S A A L T G F T F S A L Q I 394

CTGCCCTACACACTGGCCTCCCTCTACCACCGGAGAAGCAGGTGTTCTGCCCAAATAC
L P Y T L A S L Y H R E K Q V F L P K Y 414

CGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGC
R G D T G G A S S E D S L M T S F L P G 434

CCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTC
P K P G A P F P N G H V G A G G S G L L 454

CCACCTCCACCCGCGCTCTGCGGGGCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTG
P P P P A L C G A S A C D V S V R V V V 474

GGTGAGCCCACCGAGGCCAGGGTGGTTCCGGGCGGGGCATCTGCCTGGACCTCGCCATC
G E P T E A R V V P G R G I C L D L A I 494

CTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTC
L D S A F L L S Q V A P S L F M G S I V 514

CAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCCTGGGTCTGGTCGCC
Q L S Q S V T A Y M V S A A G L G L V A 534

ATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAG
I Y F A T Q V V F D K S D L A K Y S A * 554

GTCGAG

WO 03/104272

PCT/EP03/06096

22/45

FIG. 13. Sequence of the CPC-P501S expression cassette of JNW735 (SEQ ID NO:18 & 44)

Nucleotide sequence: SEQ ID NO.18

Polypeptide sequence: SEQ ID NO.44

```
#####  
GCCACCATGGCGGCGCTTACGTACATTCCGACGGCTCTTATCCAAAA  
      M A A A Y V H S D G S Y P K 14  
  
GACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTTTGACAGTTCAAGGCTATATGCTT  
D K F E K I N G T W Y Y F D S S G Y M L 34  
  
GCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAACTCAGGCGAA  
A D R W R K H T D G N W Y W F D N S G E 54  
  
ATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCC  
M A T G W K K I A D K W Y Y F N E E G A 74  
  
ATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGC  
M K T G W V K Y K D T W Y Y L D A K E G 94  
  
GCCATGCAATACATCAAGGCTAACTCTAAGTTTATTGGTATCACTGAAGGCGTCATGGTA  
A M Q Y I K A N S K F I G I T E G V M V 114  
  
TCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGA  
S N A F I Q S A D G T G W Y Y L K P D G 134  
  
ACACTGGCAGACAGGCCAGAAAAGTTCATGTACATGGTGTGTTGGGCTTGGTCCAGTGCTG  
T L A D R P E K F M Y M V L G I G P V L 154  
  
GGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGC  
G L V C V P L L G S A S D H W R G R Y G 174  
  
CGCCGCCGGCCCTTCATCTGGGCACTGTCTTGGGCATCCTGCTGAGCCTCTTTCTCATC  
R R R P F I W A L S L G I L L S L F L I 194  
  
CCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCCGATCCCAGGCCCTGGAGCTGGCA  
P R A G W L A G L L C P D P R P L E L A 214  
  
CTGCTCATCTGGGCGTGGGCTGTGGAATCTGTGGCCAGGTGTGCTTCACTCCACTG  
L L I L G V G L L D F C G Q V C F T P L 234  
  
GAGGCCCTGCTCTCTGACCTCTTCCGGGACCCGGACCACTGTGCGCAGGCCTACTCTGTC  
E A L L S D L F R D P D H C R Q A Y S V 254  
  
TATGCCTTCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGG  
Y A F M I S L G G C L G Y L L P A I D W 274  
  
GACACCAGTGCCTTGGCCCCCTACCTGGGCACCCAGGAGGTGCCTCTTTGGCCTGCTC  
D T S A L A P Y L G T Q E E C L F G L L 294  
  
ACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGCAGCGCTG  
T L I F L T C V A A T L L V A E E A A L 314  
  
GGCCCCACCGAGCCAGCAGAAGGGCTGTGCGCCCCCTCCTTGTGCCCCACTGCTGTCCA  
G P T E P A E G L S A P S L S P H C C P 334
```

WO 03/104272

PCT/EP03/06096

23/45

TGCCGGGCGCGCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTG
C R A R L A F R N L G A L L P R L H Q L 354

TGCTGCCGCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATG
C C R M P R T L R R L F V A E L C S W M 374

GCACTCATGACCTTCACGCTGTTTACACGGATTTTCGTGGGCGAGGGGCTGTACCAGGGC
A L M T F T L F Y T D F V G E G L Y Q G 394

GTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGATG
V P R A E P G T E A R R H Y D E G V R M 414

GGCAGCTGGGGCTGTTCTCGAGTGCAGCATCTCCCTGGTCTTCTCTGTGTCATGGAC
G S L G L F L Q C A I S L V F S L V M D 434

CGGCTGGTGCAGCGATTTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCT
R L V Q R F G T R A V Y L A S V A A F P 454

GTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCC
V A A G A T C L S H S V A V V T A S A A 474

CTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTAC
L T G F T F S A L Q I L P Y T L A S L Y 494

CACCGGGAGAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGT
H R E K Q V F L P K Y R G D T G G A S S 514

GAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAAT
E D S L M T S F L P G P K P G A P F P N 534

GGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCGCGCTCTGCGGGGCC
G H V G A G G S G L L P P P P A L C G A 554

TCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGGTGGTT
S A C D V S V R V V V G E P T E A R V V 574

CCGGGCGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCTGTGCTGTCCCAG
P G R G I C L D L A I L D S A F L L S Q 594

GTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTAT
V A P S L F M G S I V Q L S Q S V T A Y 614

ATGGTGTCTGCCGAGGCCTGGGTCTGGTCGCCATTTACTTTGCTACACAGGTAGTATTT
M V S A A G L G L V A I Y F A T Q V V F 634

GACAAGAGCGACTTGGCCAAATACTCAGCGTAGGTCGAG
D K S D L A K Y S A * 645

WO 03/104272

PCT/EP03/06096

24/45

FIG. 14 – Codon optimised P501S sequences (SEQ ID NO:19-20)

SEQ ID NO:19

ATGGTGCAGCGGCTCTGGGTGAGCCGCTCCTGCGGCATCGCAAGGCCAGCTCCTGCTGGTGAATCTGCTCA
CATTGCGCCTGGAGGTGTGCTGCGCGCGGCATCACCTACGTGCCCCCCTCCTGCTGGAGGTGGGAGTCGA
GGAGAAGTTCATGACCATGGTGTGCGGCATTGGGCCGCTCCTGGGCCTCGTGTGCGTGCCTCTCCTCGGCAGC
GCTTCCGACCATTTGGCGCGCGCGGTATGGCCGAGGAGACCCTTCATCTGGGCTCTGAGTCTCGGCATCCTGC
TGAGCCTGTTCTGATCCCTCGGGCCGGCTGGCTGGCCGGGCTGCTGTGCCCGGATCCTCGGCCCCCTGGAGCT
GGCCCTGCTGATCCTCGGCGTGGGCTGCTGGACTTCTGCGGCCAGGTGTGCTTACGCCCCCTGGAGGCACTG
CTGAGCGACCTGTTCCGGGACCCGACCATTTGCCGCCAGGCGTACAGCGTGTACGCCTTCATGATCTCCCTGG
GAGGCTGCCTGGGCTACCTGCTCCCCGCCATCGATTGGGACACCAGCGCACTCGCCCCCTATCTCGGAACACA
GGAGGAATGCCTGTTCCGGATTGTTGACGCTCATCTTCTCACGTGCGTGC CGGCCACCCTGTTGGTGGCCGAG
GAGGCCGCCCTGGGGCCACCGAGCCGGCCGAGGACTGAGCGCCCCGAGCCTGAGTCCACACTGCTGCCCTT
GCCGGGCCCGCTGGCCTTCCGTAATCTGGGCGCCCTCCTGCCTCGGCTCCATCAGCTGTGTTGCAGAATGCC
TAGGACGCTGCGGCGCCTGTTCTGTCGCTGAGTTGTGCTCCTGGATGGCTCTCATGACCTTCACCCTGTTTTAT
ACGGACTTCTGCGGGAGGGCCTGTACCAGGGGTGCGCGCGCCGAGCCCGGACAGAGGCGCGCCGCCACT
ACGACGAGGGAGTGCGTATGGGCTCCCTGGGCTCTTCTTGCACTGCGCCATCAGTCTGGTTTTCTCTCTGGT
CATGGACAGGCTGGTGCAGCGCTTCGGAACCCGGGCGGTGTACCTGGCGAGCGTGGCCGCTTCCCCGTGGCT
GCCGGCGCCACCTGCCTCTCTCACTCGGTGGCCGTGGTACCGCCAGCGCCGCCCTGACCGGGTTCACCTTCT
CTGCCCTGCAGATTCTGCCTTACACCCTGGCCAGCCTGTACCATCGCGAGAAACAGGTGTTTCTCCCCAAGTA
CAGAGGCGACACCGGGGCGCCTCCAGCGAGGACAGCCTCATGACCTCCTTCTGCTGGCCCCAAGCCCGGC
GCCCTTTTCCCCAACGGGCACGTGGGCGCCGGCGGAGTGGGCTCCTGCCCCCCTCCTGCGCTGTGCGGGG
CCAGCGCCTGCGACGTGAGCGTGC CGTGGTGGTGGGCGAGCCACCGAGGCCCGCGTGGTGGCGGGCAGAGG
CATTTGTCTGGACCTGGCCATCCTCGACTCCGCTTCTCTCCTCAGCCAGGTGGCCCCGTCCCTCTTCATGGGC
TCTATCGTCCAGCTGTCTCAGAGCGTCACCGCTTACATGGTGTCCGCTGCTGGACTGGGCTTGGTGGCTATTT
ATTCGCCACCCAGGTGGTGTTCGACAAGAGCGACCTGGCCAAATACTCCGCCTGA

SEQ ID NO:20

ATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGGCCAGTTGCTGCTGGTGAACCTGCTGA
CTTTCGGACTGGAGGTGTGCTGCGTGGCTGCGGGGATCACGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGA
GGAGAAGTTCATGACAATGGTGTGCGGCATCGGCCCCGCTCCTGGGCCTCGTGTGTGCGCCCTCCTCGGGAGT
GCGTCCGATCATTTGGCGGGGCGCTACGGCCGCGCAGACCGTTTCATCTGGGCCCTGAGCCTGGGGATCCTGC
TCTCTCTCTTCTGATCCCCGGGCGGCTGGCTGGCCGGCCTGCTGTGTCCCAGCCCCCGCCCTCTGGAGCT
GGCCCTCCTGATCCTGGGCGTGGGCTTGTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCTGGAGGCTCTG
CTCTCCGACCTCTTCCGCGACCCGACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGG

WO 03/104272

PCT/EP03/06096

25/45

GGGGATGCCTGGGCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTCA
GGAGGAGTGCCTGTTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAG
GAGGCGGCCCTGGGGCCACCGAGCCCGCCGAGGGCCTGAGCGCTCCCAGCCTGAGCCCCCATTGCTGCCCGT
GCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCTGCATCAGCTGTGCTGTCGCATGCC
TCGCACCTGCGCCGCTGTTTCGTCGCTGAGCTCTGTTCTGATGGCCCTGATGACGTTACCCCTCTTCTAC
ACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGCCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATT
ACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCAGTCTGGTGTCTCTCTGGT
GATGGACCGGCTGGTGCAGCGCTTCGGCACCCGGGCGGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCGCC
GCCGGCGCGACCTGCCTGTCTCATTCTGTGCGCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCA
GTGCGCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCTGCCCCAAGTA
CCGCGGGGACACAGGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCCCAAGCCGGGG
GCCCCCTTTCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCG
CTAGTGCCTGCGACGTGAGCGTGCGGGTGGTGGTGGGGGAGCCACCGAGGCTAGGGTCGTGCCTGGCCGGGG
GATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCAGGTGGCGCCAGCCTGTTTCATGGGC
AGTATCGTGAGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGCGCTGGGGTTGGTGGCCATCT
ACTTTGCCACCCAGGTCGTGTTTCGACAAGAGCGATCTCGCCAAGTATAGCGCCTGA

WO 03/104272

PCT/EP03/06096

26/45

FIG. 15 – Re-engineered codon optimised sequence 19 (SEQ ID NO:21)

GACGGCTAGCGCCACCATGGTGCAGCGGCTCTGGGTGAGCCGCTCCTGCGGCATCGCAAGGCCCAGCTCCTG
CTGGTGAATCTGCTCACATTCGGCCTGGAGGTGTGCTGGCCGCCGCATCACCTACGTGCCCCCCTCCTGC
TGGAGGTGGGAGTCGAGGAGAAGTTCATGACCATGGTGTGGGCATTGGGCCCGTCTGGGCCCTGCTGTGCGT
GCCTCTCCTCGGCAGCGCTTCCGACCATTGGCGCGGCCGTATGGCCGCAGGAGACCCCTTCATCTGGGCTCTG
AGTCTCGGCATCCTGCTGAGCCTGTTCTGATCCCTCGGGCCGGCTGGCTGGCCGGGCTGCTGTGCCCCGATC
CTCGGCCCTTGGAGCTGGCCCTGCTGATCCTCGGCGTGGCCCTGCTGGACTTCTGCGGCCAGGTGTGCTTCAC
GCCCCGAGGCACTGCTGAGCGACCTGTTCCGGGACCCGACCATTGCCGCCAGGCGTACAGCGTGTACGCC
TTCATGATCTCCCTGGGAGGCTGCCTGGGCTACCTGCTCCCGCCATCGATTGGGACACCAGCGCACTCGCCC
CCTATCTCGGAACACAGGAGGAATGCCTGTTCCGA[□]_□TGACGCTCATCTTCCTCACGTGCGTTCGGGCCAG
CCTGTTGGTGGCCGAGGAGGCCGCCCTGGGGCCACCGAGCCGGCCGAGGACTGAGCGCCCCGAGCCTGAGT
CCACACTGCTGCCCTTGCCGGGCCCGCCTGGCCTTCCGTAATCTGGGCGCCCTCCTGCCTCGGCTCCATCAGC
TGTGTTGCAGAATGCCTAGGACGCTGCGGCGCCTGTTCTGCTGAGTTGTGCTCCTGGATGGCTCTCATGAC
CTTACCCTGTTTTATACGGACTTCGTGCGGGAGGGCCTGTACCAGGGGGTGCCGCGCGCCGAGCCCGGACA
GAGGCGCGCCGCACTACGACGAGGAGTGCGTATGGGCTCCCTGGGCTCTTCTTGCACTGCGCCATCAGTC
TGGTTTTCTCTCTGGTTCATGGACAGGCTGGTGCAGCGCTTCGGAACCCGGGCGGTGTACCTGGCGAGCGTGGC
CGCCTTCCCCGTGGCTGCCGGCGCCACCTGCCTCTCTCACTCGGTGGCCGTGGTACCGCCAGCGCCGCCCTG
ACCGGGTTACACCTTCTCTGCCCTGCAGATTCTGCCTTACACCCTGGCCAGCCTGTACCATCGCGAGAAACAGG
TGTTTCTCCCAAGTACAGAGGCGACACCGGGGCGCCTCCAGCGAGGACAGCCTCATGACCTCCTTCTGCC
TGGCCCCAAGCCCGCGCCCCCTTTCCCCAACGGGCACGTGGGCGCCGGCGGAGTGGGCTCCTGCCCCCCCCT
CCTGCGCTGTGCGGGGCCAGCGCCTGCGACGTGAGCGTGCCTGCTGGTGGGCGAGCCACCGAGGCCCGCG
TGGTGCCGGGCGAGGCAATTTGCTGAGCCTGGCCATCCTCGACTCCGCCTTCTCCTCAGCCAGGTGGCCCC
GTCCCTCTTCATGGGCTCTATCGTCCAGCTGTCTCAGAGCGTCACCGCTTACATGGTGTCCGCTGCTGGACTG
GGCTTGGTGGCTATTTATTTGCCACCCAGGTGGTGTTCGACAAGAGCGACCTGGCCAAATACTCCGCCTGAC
TCGAGGCAG

WO 03/104272

PCT/EP03/06096

27/45

FIG. 16 – Re-engineered codon optimised sequence 20 (SEQ ID NO:22)

GACGGCTAGCGCCACCATGGTGACGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGGCCAGTTGCTG
CTGGTGAACTGCTGACTTTCGGACTGGAGGTGTGCCTGGCTGCCGGGATCACGTACGTGCCCCCTGCTGC
TGGAGGTGGGCGTGGAGGAGAAGTTCATGACAATGGTGCTGGGCATCGGCCCGTCCTGGGCTCGTGTGTGT
GCCCCCTCCTCGGGAGTGCCTCCGATCATTGGCGGGGCCGCTACGGCCGCGCAGACCGTTCATCTGGGCCCTG
AGCCTGGGATCCTGCTCTCTCTCTTCTCTGATCCCCGGGCGCGGCTGGCTGGCCGGCCTGTGTGTCCCAGCC
CCCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCCTGGGACTTCTGCGGCCAGGTGTGTTTCACT
TCCCCGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACGCC
TTCATGATCAGTCTGGGGGGATGCCTGGGCTATCTGCTGCCCGCTATCGACTGGGACACCAGCGCCCTGGCCC
CCTACCTGGGGAAGTACAGGAGGTGCCTGTTCCGGCCTGCTCACCTTGATCTTCTGACGTGCGTGCCTGCCAC
CCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCACCGAGCCCGCCGAGGGCCTGAGCGCTCCAGCCTGAGC
CCCCATTGCTGCCCGTGCAGGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCGCCTGCATCAGC
TGTGCTGTGCGATGCCCTCGCACCTGCGCCGCTGTTCTGCTGAGCTCTGTTCTGATGGCCCTGATGAC
GTTACCCCTCTTCTACACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGCCAGGGCCGAGCCCGGCACC
GAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCTCTTCTGTCAGTGGCCATCAGTC
TGGTGTCTCTCTGGTGTGACCGGCTGGTGACGCGCTTCGGCACCCGGGCGGTGACCTCGCCTCTGTGGC
GGCTTTCCCGTCCCGCGCGCGGCGACCTGCCTGTCTCATTTCTGTCGCGGTGGTGACCGCCAGCGCCGCCCTG
ACCGGCTTACCTTCAGTGGCTCCAGATTCTGCCCTACACCTGGCGTCTCTGTACCATCGCGAGAAGCAGG
TGTTCTGCCCCAAGTACCGCGGGGACACAGGGGAGCTTCCTCTGAGGACAGCCTGATGACCAGCTTCTTGCC
CGGCCCAAGCCGGGGGCCCTTTCCCCAACGGCCATGTCCGGGCGGGCGGCAGCGGCTGTCCCTCCCCC
CCCGCCCTGTGCGGCGTAGTGCTGCGACGTGAGCGTGGGGTGGTGGTGGGGAGCCACCGAGGCTAGGG
TCGTGCCTGGCCGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCTGCTCTCCAGGTGGCGCC
CAGCCTGTTTCATGGGCAGTATCGTGACGTGAGCCAGAGCGTGACCGCTACATGGTGAGCGCGCGCGGCTG
GGGTGGTGGCCATCTACTTTGCCACCCAGGTGCTGTTGACAAGAGCGATCTCGCCAAGTATAGCGCCTGAC
TCGAGGCAG

WO 03/104272

PCT/EP03/06096

28/45

FIG. 17 – The starting sequence for the optimisation of CPC (SEQ ID NO:23)

Four amino acids of P501S sequence are boxed.

ATGGCGGCCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGT
ACTACTTTGACAGTTTCAAGGCTATATGCTTGACAGCCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTT
CGACAACCTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGT
GCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGCAAT
ACATCAAGGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTTTATCCAGTCAGC
GGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAAGTTCATGTAC

FIG. 18 – Representative codon optimised CPC sequences (SEQ ID NO:24-25)

SEQ ID NO:24

ATGGCGGCCGCTTACGTGCATAGCGACGGGAGCTACCCCAAGGACAAGTTTCGAGAAGATCAACGGGACATGGT
ACTACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTT
CGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGC
GCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCATGCAGT
ATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCAACGCCTTTATCCAGAGCGC
CGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCTCGCGGATCGGCCCGAGAAGTTCATGTAC

SEQ ID NO:25

ATGGCGGCCGCTTACGTGCACAGCGACGGGTCTACCCAAAGGACAAGTTTCGAGAAGATCAACGGCACGTGGT
ACTATTTTCGACAGCAGCGCTACATGCTCGCCGATCGCTGGCGCAAGCACACCGACGGGAAGTGGTACTGGTT
CGACAACCTCGGCGAGATGGCTACGGGTGGAAGAAGATCGCGGACAAGTGGTACTACTTCAACGAGGAGGGC
GCCATGAAGACCGGCTGGGTGAAGTACAAGGACACCTGGTACTACCTGGACGCTAAGGAGGGCGCCATGCAGT
ACATCAAGGCCAACTCGAAGTTCATCGGGATCACCGAGGGCGTGATGGTCAGTAACGCTTTATCCAGAGCGC
GGACGGCACAGGCTGGTATTACCTGAAGCCCGATGGCACCTGGCGGACAGACCTGAGAAATTTCATGTAC

FIG. 19 – Engineered CPC codon optimised sequence (SEQ ID NO:26)

SEQ ID NO:26

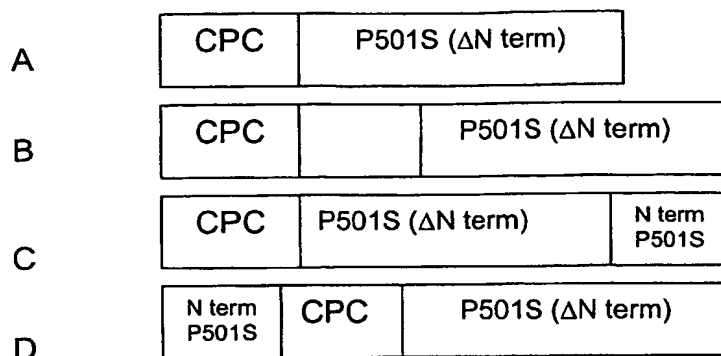
GACGGCTAGCGCCACCATGGCCGCCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACAAGTTTCGAGAAG
ATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACG
GCAACTGGTACTGGTTCGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTA
TTTCAACGAGGAGGGCGCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAG
GAGGGCGCCATGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCAACG
CCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCTCGCGGATCGGCCGGA
AAGTTCATGTACTGACTCGAGGCAG

WO 03/104272

PCT/EP03/06096

29/45

FIG. 20 – P501S CPC fusion candidate constructs and sequences



Construct A = SEQ ID NO:37 (nucleotide) & 45 (polypeptide)

GCGGCCGCGCCACCATGGCCGCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
 M A A A Y V H S D G S Y P K D K
 AGTTCGAGAAGATCAACGGGACATGGTACTACTTTCGACTCCTCCGGCTACATGTCGCCG
 F E K I N G T W Y Y F D S S G Y M L A D
 ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
 R W R K H T D G N W Y W F D N S G E M A
 CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCACGAGGAGGGCGCCATGA
 T G W K K I A D K W Y Y F N E E G A M K
 AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
 T G W V K Y K D T W Y Y L D A K E G A M
 TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
 Q Y I K A N S K F I G I T E G V M V S N
 ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
 A F I Q S A D G T G W Y Y L K P D G T L
 TCGCGGATCGGCCCCGAGAAGTTCATGTACATGGTGTCTGGGCATCGGCCCCGTCCTGGGCC
 A D R P E K F M Y M V L G I G P V L G L
 TCGTGTGTGTGCCCCCTCCTCGGGAGTGCGTCCGATCATTGGCGGGGCGGCTACGGCCGCC
 V C V P L L G S A S D H W R G R Y G R R
 GCAGACCGTTTCATCTGGGCCCTGAGCCTGGGCATCCTGCTCTCTCTCTCTCTCTGATCCCC
 R P F I W A L S L G I L L S L F L I P R
 GGGCCGGCTGGCTGGCCGGCCTGTGTGTCCCGACCCCGCCCTCTGGAGCTGGCCCTCC
 A G W L A G L L C P D P R P L E L A L L
 TGATCCTGGGCGTGGGCCTGCTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGG

WO 03/104272

PCT/EP03/06096

30/45

I L G V G L L D F C G Q V C F T P L E A
CTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGATAC
L L S D L F R D P D H C R Q A Y S V Y A
CCTTCATGATCAGTCTGGGGGATGCCTGGGCTATCTGCTGCCCCGCTATCGACTGGGACA
F M I S L G G C L G Y L L P A I D W D T
CCAGCGCCCTGGCCCCCTACCTGGGGACTCAGGAGGAGTGCCCTGTTCCGGCCTGCTCACCT
S A L A P Y L G T Q E E C L F G L L T L
TGATCTTCCTGACGTGCGTCGCCGCCACCCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGC
I F L T C V A A T L L V A E E A A L G P
CCACCGAGCCCGCGAGGGCCTGAGCGCTCCAGCCTGAGCCCCATTGCTGCCCGTGCA
T E P A E G L S A P S L S P H C C P C R
GGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCCTGCATCAGCTGTGCT
A R L A F R N L G A L L P R L H Q L C C
GTCGATGCTCGCACCCCTGCGCCGCTGTTGCTGCTGAGCTCTGTTCTGATGGCCC
R M P R T L R R L F V A E L C S W M A L
TGATGACGTTACCCCTCTTCTACACCGACTTCGTGGGGAGGGCCTGTACCAGGGCGTGC
M T F T L F Y T D F V G E G L Y Q G V P
CCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCT
R A E P G T E A R R H Y D E G V R M G S
CTCTGGGCTCTTCTGCGAGTGCGCCATCAGTCTGGTGTCTCTCTGATGGACCGGC
L G L F L Q C A I S L V F S L V M D R L
TGGTGCAGCGCTTCGGCACCCGGGCGGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCG
V Q R F G T R A V Y L A S V A A F P V A
CCGCCGGCGCGACCTGCCTGTCTCATTCTGTGCGCGTGGTGACCGCCAGCGCCGCCCTGA
A G A T C L S H S V A V V T A S A A L T
CCGGCTTCACCTTCAGTGCCTCCAGATTCTGCCCTACACCTGGCGTCTCTGTACCATC
G F T F S A L Q I L P Y T L A S L Y H R
GCGAGAAGCAGGTGTTCTGCCCCAAGTACCGCGGGGACACAGGGGGAGCTTCCTCTGAGG
E K Q V F L P K Y R G D T G G A S S E D
ACAGCCTGATGACCACTTCTTGCCCGGCCCAAGCGGGGGCCCCCTTTCCCCAACGGCC
S L M T S F L P G P K P G A P F P N G H
ATGTCGGGGCGGGCGGCGAGCGGCTGCTCCCTCCCCCCCCCGCCTGTGCGGCGCTAGTG
V G A G G S G L L P P P P A L C G A S A
CCTGCGACGTGAGCGTGGGGTGGTGGTGGGGAGCCACCGAGGCTAGGGTCTGCGCTG
C D V S V R V V V G E P T E A R V V P G
GCCGGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCCTTCCTGCTCTCCAGGTGG
R G I C L D L A I L D S A F L L S Q V A
CGCCAGCCTGTTTCATGGGCAGTATCGTGCAGCTGAGCCAGAGCGTGACCGCCTACATGG
P S L F M G S I V Q L S Q S V T A Y M V

WO 03/104272

PCT/EP03/06096

31/45

TGAGCGCCGCCCGCCTGGGGTTGGTGGCCATCTACTTTGCCACCCAGGTCGTGTTTCGACA
S A A G L G L V A I Y F A T Q V V F D K

AGAGCGATCTCGCCAAGTATAGCGCCTGAGGATCC
S D L A K Y S A *

Construct B = SEQ ID NO:38 (nucleotide) & 46 (polypeptide)

GCGGCCGCGCCACCATGGCCGCCGCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
M A A A Y V H S D G S Y P K D K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
F E K I N G T W Y Y F D S S G Y M L A D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
R W R K H T D G N W Y W F D N S G E M A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
T G W K K I A D K W Y Y F N E E G A M K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
T G W V K Y K D T W Y Y L D A K E G A M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
Q Y I K A N S K F I G I T E G V M V S N

ACGCCTTTATCCAGAGCGCCGACGGCACC GGATGGTACTACTTGAAGCCGGACGGCACCC
A F I Q S A D G T G W Y Y L K P D G T L

TCGCGGATCGGCCCCGAGATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAA
A D R P E M V Q R L W V S R L L R H R K

AGGCCCAAGTTGCTGCTGGTGAACCTGCTGACTTTCGACTGGAGGTGTGCTGGCTGCCG
A Q L L L V N L L T F G L E V C L A A G

GGATCACGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGAAGTTTCATGACAA
I T Y V P P L L L E V G V E E K F M T M

TGGTGCTGGGCATCGGCCCCGCTCGTGGGCTCGTGTGTGCCCCCTCCTCGGGAGTGCCT
V L G I G P V L G L V C V P L L G S A S

CCGATCATTTGGCGGGGCCGCTACGGCCGCCGACCGTTCATCTGGGCCCTGAGCCTGG
D H W R G R Y G R R R P F I W A L S L G

GCATCCTGCTCTCTCTCTCTCTGATCCCCGGGCCGGCTGGCTGGCCGGCCTGCTGTGTC
I L L S L F L I P R A G W L A G L L C P

CCGACCCCCGCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCTGCTGGACTTCT
D P R P L E L A L L I L G V G L L D F C

GCGGCCAGGTGTGTTTCACTCCCCTGGAGGCTCTGCTCTCCGACCTCTCCGCGACCCCG
G Q V C F T P L E A L L S D L F R D P D

ACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGGGGGATGCCTGG
H C R Q A Y S V Y A F M I S L G G C L G

GCTATCTGCTGCCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTC
Y L L P A I D W D T S A L A P Y L G T Q

WO 03/104272

PCT/EP03/06096

32/45

AGGAGGAGTGCCTGTTGCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCC
 E E C L F G L L T L I F L T C V A A T L
 TGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCACCGAGCCCGCCGAGGGCCTGAGCGCTC
 L V A E E A A L G P T E P A E G L S A P
 CCAGCCTGAGCCCCATTGCTGCCCGTGACGGGCTAGGCTCGCCTTCAGGAATCTGGGCG
 S L S P H C C P C R A R L A F R N L G A
 CTTTGCTGCCCCGCCTGCATCAGCTGTGCTGTCGCATGCCTCGCACCTGCGCCGCTGT
 L L P R L H Q L C C R M P R T L R R L F
 TCGTCGCTGAGCTCTGTTCTGGATGGCCCTGATGACGTTACCTCTTCTACACCGACT
 V A E L C S W M A L M T F T L F Y T D F
 TCGTGGGGGAGGGCCTGTACCAGGGCGTGCCAGGGCCGAGCCCGGCACCGAGGCTAGGC
 V G E G L Y Q G V P R A E P G T E A R R
 GCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCA
 H Y D E G V R M G S L G L F L Q C A I S
 GTCTGGTGTCTCTCTGGTGATGGACCGGCTGGTGACGCTTCGGCACCCGGGCGCTGT
 L V F S L V M D R L V Q R F G T R A V Y
 ACCTCGCCTCTGTGGCGGCTTTCCCCGTCGCCGCCGGCGCGACCTGCCTGTCTCATCTG
 L A S V A A F P V A A G A T C L S H S V
 TCGCCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCAGTGCGCTCCAGATTC
 A V V T A S A A L T G F T F S A L Q I L
 TGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCTGCCCCAAGTACC
 P Y T L A S L Y H R E K Q V F L P K Y R
 GCGGGGACACAGGGGGAGCTTCTCTGAGGACAGCCTGATGACAGCTTCTTGCCCGGCC
 G D T G G A S S E D S L M T S F L P G P
 CCAAGCCGGGGGCCCTTTCCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCC
 K P G A P F P N G H V G A G G S G L L P
 CTCCCCCCCCCGCCCTGTGCGGCGCTAGTGCTGCGACGTGAGCGTGCGGGTGGTGGTGG
 P P P A L C G A S A C D V S V R V V V G
 GGGAGCCACCGAGGCTAGGGTCGTGCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCC
 E P T E A R V V P G R G I C L D L A I L
 TCGACTCCGCCTTCCTGCTCTCCAGGTGGCGCCAGCCTGTTTCATGGGCAGTATCGTGC
 D S A F L L S Q V A P S L F M G S I V Q
 AGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGGCCTGGGGTGGTGGCCA
 L S Q S V T A Y M V S A A G L G L V A I
 TCTACTTTGCCACCCAGGTGTTGACACAAGAGCGATCTCGCCAAGTATAGCGCCTGAG
 Y F A T Q V V F D K S D L A K Y S A *

GATCC

Construct C = SEQ ID NO:39 (nucleotide) & 47 (polypeptide)

WO 03/104272

PCT/EP03/06096

33/45

GCGGCCGCGCCACCATGGCCGCGCCTACGTGCATAGCGACGGGAGCTACCCCAAGGACA
M A A A Y V H S D G S Y P K D K

AGTTCGAGAAGATCAACGGGACATGGTACTACTTCGACTCCTCCGGCTACATGCTCGCCG
F E K I N G T W Y Y F D S S G Y M L A D

ACCGCTGGCGGAAGCACACCGACGGCAACTGGTACTGGTTCGATAACTCGGGAGAGATGG
R W R K H T D G N W Y W F D N S G E M A

CCACCGGCTGGAAGAAGATCGCGGACAAGTGGTACTATTTCAACGAGGAGGGCGCCATGA
T G W K K I A D K W Y Y F N E E G A M K

AGACCGGCTGGGTGAAGTATAAGGACACCTGGTACTACCTCGACGCCAAGGAGGGCGCCA
T G W V K Y K D T W Y Y L D A K E G A M

TGCAGTATATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAGGGAGTGATGGTCAGCA
Q Y I K A N S K F I G I T E G V M V S N

ACGCCTTTATCCAGAGCGCCGACGGCACCGGATGGTACTACTTGAAGCCGGACGGCACCC
A F I Q S A D G T G W Y Y L K P D G T L

TCGCGGATCGGCCCCGAGAAGTTCATGTACATGGTGCTGGGCATCGGCCCCGTCCTGGGCC
A D R P E K F M Y M V L G I G P V L G L

TCGTGTGTGTGCCCCCTCCTCGGGAGTGCCTCCGATCATTGGCGGGGCGCTACGGCGGCC
V C V P L L G S A S D H W R G R Y G R R

GCAGACCGTTTCATCTGGGCCCTGAGCCTGGGCATCCTGCTCTCTCTCTCTCTGATCCCC
R P F I W A L S L G I L L S L F L I P R

GGGCGGCTGGCTGGCCGCGCTGCTGTGTCCCGACCCCGCCCTCTGGAGCTGGCCCTCC
A G W L A G L L C P D P R P L E L A L L

TGATCCTGGGCGTGGCCCTGCTGGACTTCTGCGGCCAGGTGTGTTTCACTCCCCTGGAGG
I L G V G L L D F C G Q V C F T P L E A

CTCTGCTCTCCGACCTCTTCCGCGACCCCGACCACTGTAGGCAGGCTTACAGCGTGTACG
L L S D L F R D P D H C R Q A Y S V Y A

CCTTCATGATCAGTCTGGGGGATGCCTGGGCTATCTGCTGCCCCGCTATCGACTGGGACA
F M I S L G G C L G Y L L P A I D W D T

CCAGCGCCCTGGCCCCCTACCTGGGGACTCAGGAGGAGTGCCTGTTCCGGCTGCTCACCT
S A L A P Y L G T Q E E C L F G L L T L

TGATCTTCTGACGTGCGTCGCCGCCACCTGCTGGTGGCCGAGGAGGCGGCCCTGGGGC
I F L T C V A A T L L V A E E A A L G P

CCACCGAGCCCGCGAGGGCTGAGCGCTCCCAGCCTGAGCCCCATTGCTGCCCCGTGCA
T E P A E G L S A P S L S P H C C P C R

GGGCTAGGCTCGCCTTCAGGAATCTGGGCGCTTTGCTGCCCCGCTGCATCAGCTGTGCT
A R L A F R N L G A L L P R L H Q L C C

GTCGCATGCCTCGCACCTGCGCCGCTGTTGCTGCTGAGCTCTGTTCTGGATGGCCC
R M P R T L R R L F V A E L C S W M A L

TGATGACGTTACCCCTCTTCTACACCGACTTCGTGGGGGAGGGCCTGTACCAGGGCGTGC
M T F T L F Y T D F V G E G L Y Q G V P

WO 03/104272

PCT/EP03/06096

34/45

CCAGGGCCGAGCCCGGCACCGAGGCTAGGCGCCATTACGACGAGGGCGTCAGGATGGGCT
R A E P G T E A R R H Y D E G V R M G S
CTCTGGGCCTCTTCTGTCAGTGCGCCATCAGTCTGGTGTTCCTCTCTGGTGATGGACCGGC
L G L F L Q C A I S L V F S L V M D R L
TGGTGCAGCGCTTCGGCACC CGGGCCGTGTACCTCGCCTCTGTGGCGGCTTTCCCCGTCG
V Q R F G T R A V Y L A S V A A F P V A
CCGCCGCGCGACCTGCCTGTCTCATTCTGTGCGCGTGGTGACCGCCAGCGCGCCCTGA
A G A T C L S H S V A V V T A S A A L T
CCGGCTTACCTTCAGTGCCTCCAGATTCTGCCCTACACCCTGGCGTCTCTGTACCATC
G F T F S A L Q I L P Y T L A S L Y H R
GCGAGAAGCAGGTGTTCTGCCCCAAGTACCGCGGGGACACAGGGGGAGCTTCTCTGAGG
E K Q V F L P K Y R G D T G G A S S E D
ACAGCCTGATGACCAGCTTCTTGCCCCGCCCAAGCGGGGGCCCCCTTTCCCCAACGGCC
S L M T S F L P G P K P G A P F P N G H
ATGTCGGGGCGGGCGGCAGCGGCCTGCTCCCTCCCCCCCCCGCCCTGTGCGGCGTAGTG
V G A G G S G L L P P P P A L C G A S A
CCTGCGACGTGAGCGTCCGGGTGGTGGTGGGGAGCCACCGAGGCTAGGGTCTGTCCTG
C D V S V R V V V G E P T E A R V V P G
GCCCGGGGATCTGCCTGGACCTGGCCATCCTCGACTCCGCTTCTGCTCTCCAGGTGG
R G I C L D L A I L D S A F L L S Q V A
CGCCAGCCTGTTTCATGGGCAGTATCGTGCAGCTGAGCCAGAGCGTGACCGCTACATGG
P S L F M G S I V Q L S Q S V T A Y M V
TGAGCGCGCGCGCCTGGGGTGGTGGCCATCTACTTTGCCACCCAGGTGCTGTTGACA
S A A G L G L V A I Y F A T Q V V F D K
AGAGCGATCTCGCCAAGTATAGCGCCATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGC
S D L A K Y S A M V Q R L W V S R L L R
GCCATAGAAAGGCCAGTTGCTGCTGGTGAACCTGCTGACTTTCGGAAGTGGAGGTGTGCC
H R K A Q L L L V N L L T F G L E V C L
TGGCTGCCGGGATCACGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGTGAG
A A G I T Y V P P L L L E V G V E E *
GATCC

Construct D = SEQ ID NO:40 (nucleotide) & 48 (polypeptide)

GCGGCCGCGCCACCATGGTGCAGCGGCTGTGGGTGTCCCGGCTGCTGCGCCATAGAAAGG
M V Q R L W V S R L L R H R K A
CCCAGTTGCTGCTGGTGAACCTGCTGACTTTCGGAAGTGGAGGTGTGCTGGCTGCCGGGA
Q L L L V N L L T F G L E V C L A A G I
TCACGTACGTGCCCCCCTGCTGCTGGAGGTGGGCGTGGAGGAGATGGCCGCGCCTACG
T Y V P P L L L E V G V E E M A A A Y V

WO 03/104272

PCT/EP03/06096

35/45

TGCATAGCGACGGGAGCTACCCCAAGGACAAGTTCGAGAAGATCAACGGGACATGGTACT
H S D G S Y P K D K F E K I N G T W Y Y
ACTTCGACTCCTCCGGCTACATGCTCGCCGACCGCTGGCGGAAGCACACCGACGGCAACT
F D S S G Y M L A D R W R K H T D G N W
GGTACTGGTTCGATAACTCGGGAGAGATGGCCACCGGCTGGAAGAAGATCGCGGACAAGT
Y W F D N S G E M A T G W K K I A D K W
GGTACTATTTCAACGAGGAGGGCGCCATGAAGACCGGCTGGGTGAAGTATAAGGACACCT
Y Y F N E E G A M K T G W V K Y K D T W
GGTACTACCTCGACGCCAAGGAGGGCGCCATGCAGTATATCAAGGCCAACAGCAAGTTCA
Y Y L D A K E G A M Q Y I K A N S K F I
TCGGCATCACCGAGGGAGTGATGGTCAGCAACGCCTTTATCCAGAGCGCCGACGGCACCG
G I T E G V M V S N A F I Q S A D G T G
GATGGTACTACTTGAAGCCGGACGGCACCCCTCGCGGATCGGCCCCGAGAAGTTTCATGTACA
W Y Y L K P D G T L A D R P E K F M Y M
TGGTGCTGGGCATCGGCCCCGTCCTGGGCCTCGTGTGTGCCCCCTCCTCGGGAGTGCCT
V L G I G P V L G L V C V P L L G S A S
CCGATCATTTGGCGGGGCGCTACGGCCGCGCAGACCGTTTCATCTGGGCCCTGAGCCTGG
D H W R G R Y G R R R P F I W A L S L G
GCATCCTGCTCTCTCTCTCTCTGATCCCCGGGCGGCTGGCTGGCCGCGCTGCTGTGTC
I L L S L F L I P R A G W L A G L L C P
CCGACCCCCGCCCCCTCTGGAGCTGGCCCTCCTGATCCTGGGCGTGGGCCTGCTGGACTTCT
D P R P L E L A L L I L G V G L L D F C
CGCGCCAGGTGTGTTTCACTCCCCCTGGAGGCTCTGCTCTCCGACCTCTTCCGCGACCCCG
G Q V C F T P L E A L L S D L F R D P D
ACCACTGTAGGCAGGCTTACAGCGTGTACGCCTTCATGATCAGTCTGGGGGGATGCCTGG
H C R Q A Y S V Y A F M I S L G G C L G
GCTATCTGCTGCCCCGCTATCGACTGGGACACCAGCGCCCTGGCCCCCTACCTGGGGACTC
Y L L P A I D W D T S A L A P Y L G T Q
AGGAGGAGTGCCTGTTCGGCCTGCTCACCTTGATCTTCCTGACGTGCGTCGCCGCCACCC
E E C L F G L L T L I F L T C V A A T L
TGCTGGTGGCCGAGGAGGCGGCCCTGGGGCCACCGAGCCCCGAGGGCCTGAGCGCTC
L V A E E A A L G P T E P A E G L S A P
CCAGCCTGAGCCCCATTGCTGCCCCGTGACGGGCTAGGCTCGCCTTCAGGAATCTGGGCG
S L S P H C C P C R A R L A F R N L G A
CTTTGCTGCCCCGCTGTCATCAGCTGTGCTGTGTCGATGCCTCGCACCTGCGCCGCTGT
L L P R L H Q L C C R M P R T L R R L F
TCGTCGCTGAGCTCTGTTCTGGATGGCCCTGATGACGTTACCCCTCTTCTACACCGACT
V A E L C S W M A L M T F T L F Y T D F
TCGTGGGGGAGGGCCTGTACCAGGGCGTCCCAGGGCCGAGCCCGGACCGAGGCTAGGC

WO 03/104272

PCT/EP03/06096

36/45

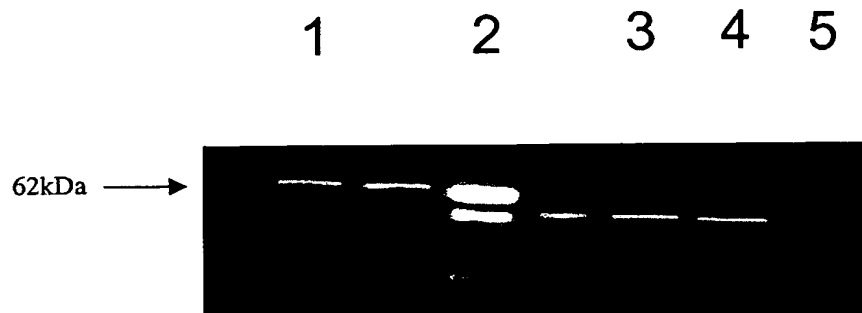
V G E G L Y Q G V P R A E P G T E A R R
GCCATTACGACGAGGGCGTCAGGATGGGCTCTCTGGGCCTCTTCCTGCAGTGCGCCATCA
H Y D E G V R M G S L G L F L Q C A I S
GTCTGGTGTCTCTCTGGTGATGGACCGCTGGTGCAGCGCTTCGGCACCCGGGCCGTGT
L V F S L V M D R L V Q R F G T R A V Y
ACCTCGCCTCTGTGGCGGCTTTCCCCGTCGCCGCCGGCGCGACCTGCCTGTCTCATTCTG
L A S V A A F P V A A G A T C L S H S V
TCGCCGTGGTGACCGCCAGCGCCGCCCTGACCGGCTTCACCTTCAGTGCGCTCCAGATTCC
A V V T A S A A L T G F T F S A L Q I L
TGCCCTACACCCTGGCGTCTCTGTACCATCGCGAGAAGCAGGTGTTCTGCCCAAGTACC
P Y T L A S L Y H R E K Q V F L P K Y R
CGGGGACACAGGGGAGCTTCTCTGAGGACAGCCTGATGACCAGCTTCTTGCCCGGCC
G D T G G A S S E D S L M T S F L P G P
CCAAGCCGGGGGCCCTTTCCCCAACGGCCATGTCGGGGCGGGCGGCAGCGGCCTGCTCC
K P G A P F P N G H V G A G G S G L L P
CTCCCCCCCCCGCCCTGTGCGCGCTAGTGCCCTGCGACGTGAGCGTGCGGGTGGTGGTGG
P P P A L C G A S A C D V S V R V V V G
GGGAGCCACCGAGGCTAGGGTCGTGCCTGGCCGGGGGATCTGCCTGGACCTGGCCATCC
E P T E A R V V P G R G I C L D L A I L
TCGACTCCGCCTTCCTGCTCTCCAGGTGGCGCCAGCCTGTTTCATGGGCAGTATCGTGC
D S A F L L S Q V A P S L F M G S I V Q
AGCTGAGCCAGAGCGTGACCGCCTACATGGTGAGCGCCGCCGCTGGGGTTGGTGGCCA
L S Q S V T A Y M V S A A G L G L V A I
TCTACTTTGCCACCCAGGTCGTGTTGACAAAGAGCGATCTCGCCAAGTATAGCGCCTGAG
Y F A T Q V V F D K S D L A K Y S A *
GATCC

WO 03/104272

PCT/EP03/06096

37/45

FIG. 21 – Western blot analysis of CHO cells following transient transfection with P501S (JNW680), CPC-P501S (JNW735) and empty vector control.



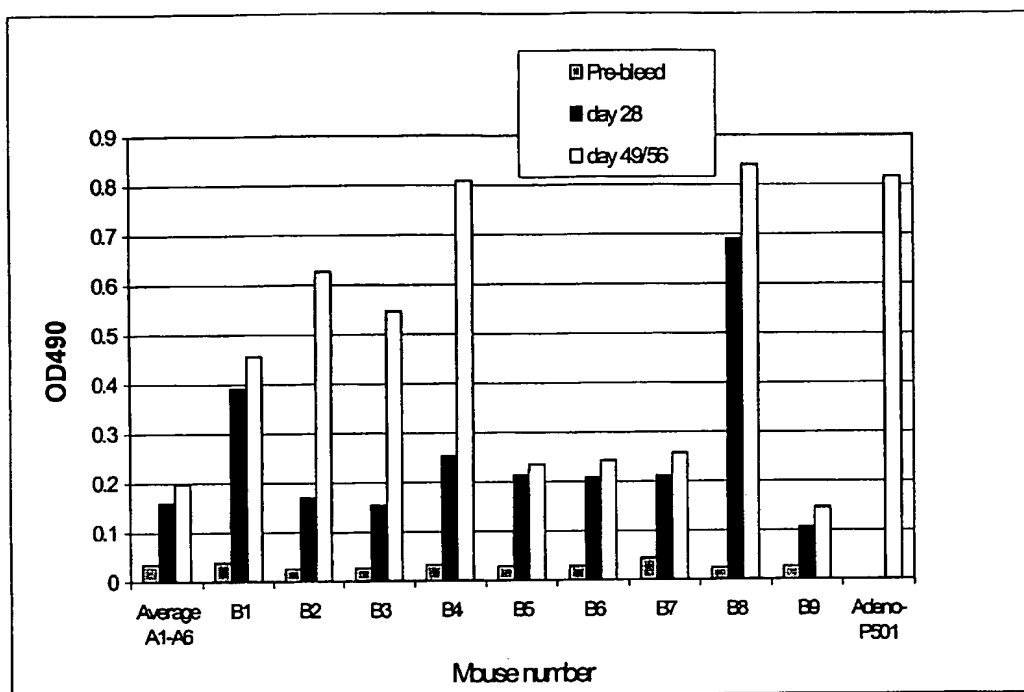
Lane	Sample
1	CPC-P501S (JNW735)
2	CPC P501S protein (62.5ng)
3	P501S (JNW680)
4	P501S (JNW680)
5	Empty vector control

WO 03/104272

PCT/EP03/06096

38/45

FIG. 22 – Anti-P501S antibody responses following immunisation at day0, 21 & 42 with pVAC-P501S (JNW680, mice B1-9) or Empty vector (pVAC, mice A1-6).

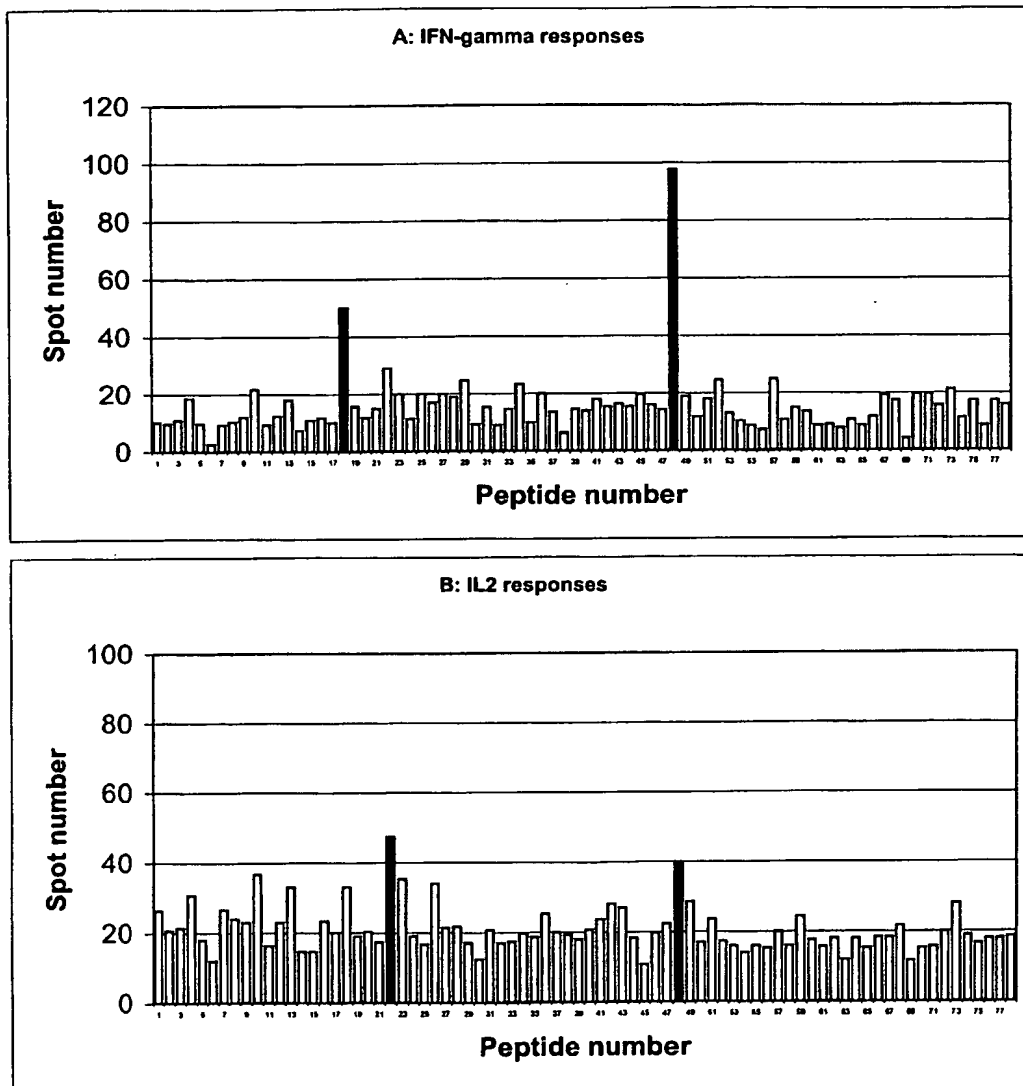


WO 03/104272

PCT/EP03/06096

39/45

FIG. 23 – Peptide library screen using C57BL/6 mice immunised at day 0, 21, 42, and 70 with pVAC-P501S (JNW680).



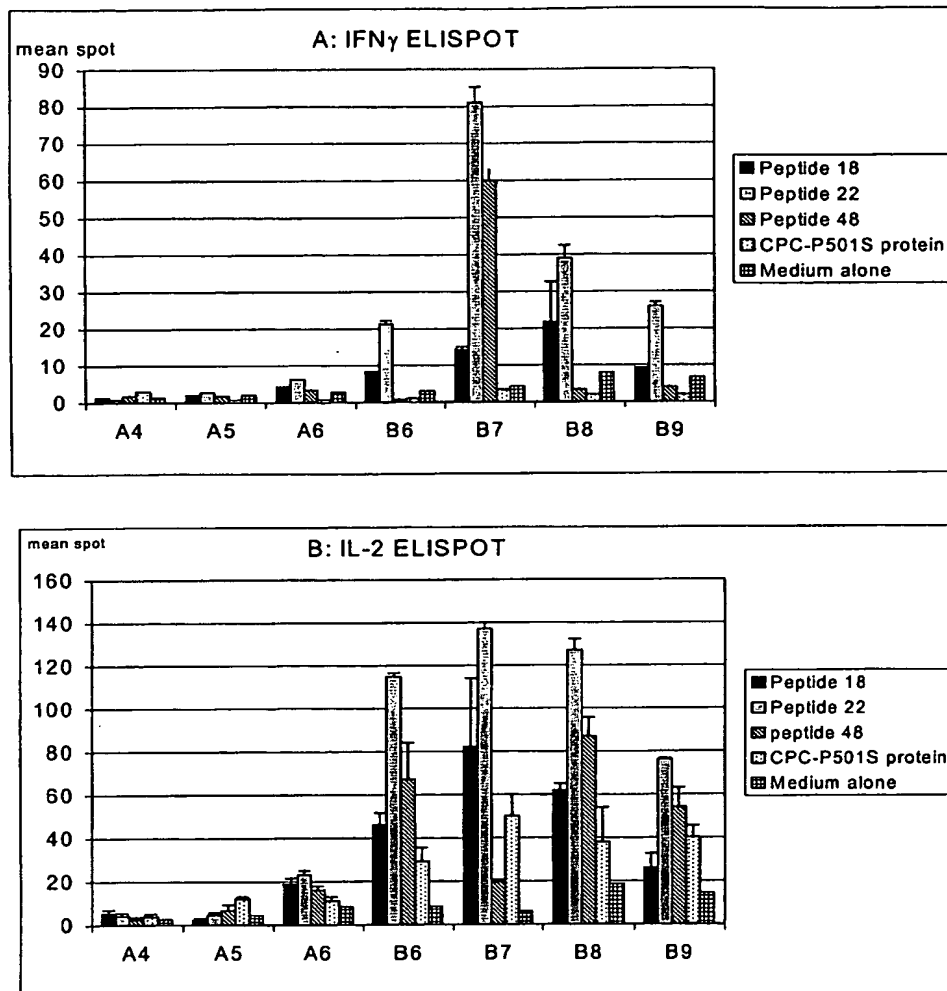
WO 03/104272

PCT/EP03/06096

40/45

FIG. 24 – Cellular responses by ELISPOT at day 77 following PMID immunisation at day 0, 21, 42, and 70 with pVAC-P501S (JNW680, B6-9) and pVAC empty (A4-6).

Graph A shows the IFN- γ responses whilst Graph B shows the IL-2 responses.



WO 03/104272

PCT/EP03/06096

41/45

FIG. 25 – Comparison of P501S and CPC-P501S.

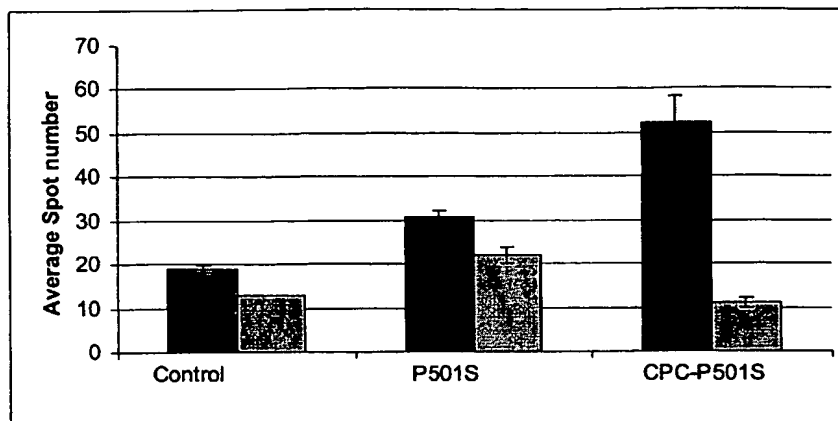
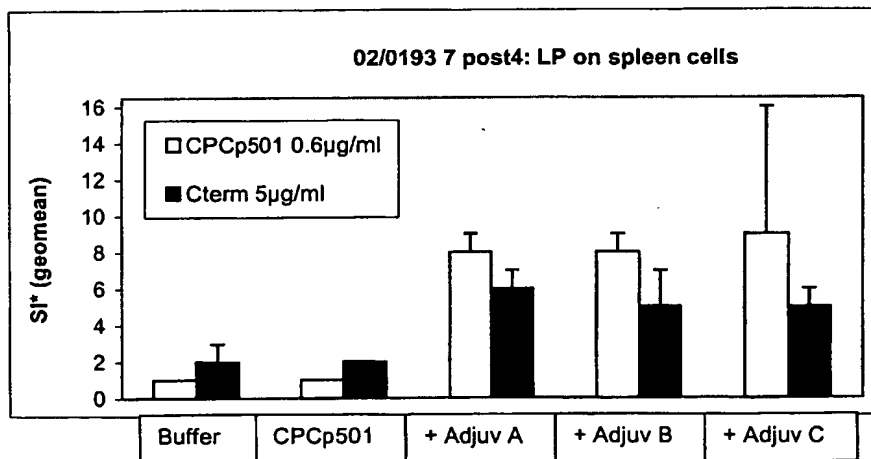


FIG. 26 – Immune response (lymphoproliferation on spleen cells) following protein immunisation with CPC-P501S.

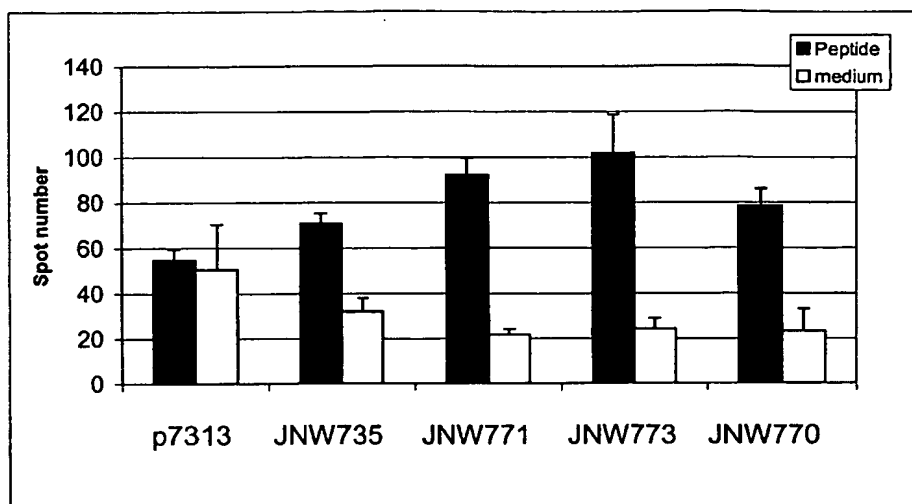


WO 03/104272

PCT/EP03/06096

42/45

FIG. 27 – Evaluation of the immune response to different CPC-P501S constructs



WO 03/104272

PCT/EP03/06096

43/45

FIG.28. MUC1-CPC DNA and polypeptide sequences

FIG. 28A. DNA sequence (SEQ ID NO.49)

ATGACACCGGGCACCCAGTCTCCTTTCTTCTGCTGCTGCTCCTCACAGTGCTTACAGTTGTTACAGGTTCTG
GTCATGCAAGCTCTACCCAGGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCAGCTCTAC
TGAGAAGAATGCTGTGAGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACT
CAGGGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAGCTTCAGGTTAGCTGCCACCTGGGGACAGGATG
TCACCTCGGTCCAGTACCAGGCCAGCCTGGGCTCCACCACCCCGCCAGCCACGATGTACCTCAGCCCC
GGACAACAAGCCAGCCCCGGGCTCCACGCCCCCCCCAGCCACGGTGTACCTCGGCCCGGACACCAGGCCG
CCCCCGGGCTCCACGCCCCCCCCAGCCACGGTGTACCTCGGCCCGGACACCAGGCCGCCCCCGGGCTCCA
CCGCGCCCGCAGCCACGGTGTACCTCGGCCCGGACACCAGGCCGCCCCCGGGCTCCACGCCCCCCCCAGC
CCATGGTGTACCTCGGCCCGGACACAGGCCCGCCTTGGCGTCCACGCCCCCTCAGTCCACAATGTACAC
TCGGCCTCAGGCTCTGCATCAGGCTCAGCTTCTACTCTGGTGCACAACGGCACCTCTGCCAGGGTACCACAA
CCCCAGCCAGCAAGAGCACTCCATTCTCAATTCAGCCACCACTCTGATACTCTACCACCTTGGCAGCCA
TAGACCAAGACTGATGCCAGTAGCACTCACCATAGCACGGTACCTCCTCTCACCTCCTCCAATCACAGCACT
TCTCCCCAGTTGTCTACTGGGGTCTCTTTCTTTTCTGCTCTTTTACATTTCAAACCTCCAGTTTAATTCTT
CTCTGGAAGATCCCAGCACCGACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTGCGAGATTTA
TAAACAAGGGGGTTTTCTGGGCCTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTG
GCCTTCCGAGAAGGTACCATCAATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACGGAAGCAGCCT
CTCGATATAACCTGACGATCTCAGACGTGAGCGTGAGTGATGTGCCATTTCTTTCTCTGCCCAGTCTGGGGC
TGGGGTGCCAGGCTGGGGCATCGCGCTGCTGGTGTGGTCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTC
ATTGCCTTGGCTGTCTGTCACTGCCGCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCCGGGATACCT
ACCATCTATGAGCGAGTACCCACCTACCACACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAG
CCCCATATGAGAAGGTTTCTGCAGGTAATGGTGGCAGCAGCCTCTCTTACACAAACCCAGCAGTGGCAGCCACT
TCTGCCAACTTGATGGCGGCCGCTTACGTACATTCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCA
ATGGCACTTGGTACTACTTTGACAGTTCAGGCTATATGCTTGACAGCCGCTGGAGGAAGCACACAGACGGCAA
CTGGTACTGGTTCGACAACCTCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTT
AACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAG
GCGCCATGCAATACATCAAGGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTT
TATCCAGTCAGCGGACGGAACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAATGA

FIG. 28B. MUC1-CPC polypeptide sequence (SEQ ID NO.50)

MTPGTQSPFFLLLLLVLTVVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMSSVLSSHSPGSGSSSTT
QGQDVT LAPATEPASGSAATWQDVTSPVTRPALGSTTPPAHDVTSAPDNKPAPGSTAPPAHGVT SAPDTRP
PPGSTAPPAHGVT SAPDTRPPPGSTAPAAHGVT SAPDTRPAPGSTAPPAHGVT SAPDNRPALASTAPPVHNVT
SASGSASGSASTLVHNGTSARATTTASKSTPFSIPSHSDTPTTLASHSTKT DASSTHHSTVPLTSSNHST
SPQLSTGVSFFFLSFHISNLQFNSSLEDPSTDYYQELQRDISEMFLQIYKQGGFLGLSNIKFRPGSVVVQLTL

WO 03/104272

PCT/EP03/06096

44/45

AFREGTINVHDTVETQFNQYKTEAASRYNLTISDVSVSDVPFPFSAQSGAGVPGWGIALLLVLCVLVALAIVYL
IALAVCQCRKKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSDRSPYEKVSAGNGGSSLSYTNPAVAAT
SANLMAAAVYHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDDNSGEMATGWKKIADKWYYF
NEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADGTGWYYLKPdGTLADRPE

FIG.29. ss-CPC-MUC1 construct and sequence

5 FIG. 29A. DNA sequence (SEQ ID NO.51)

ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGTCCACTCCCAGGTCCAAATGGCGG
CCGCTTACGTACATTCCGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTTGGTACTACTT
TGACAGTTTACAGGCTATATGCTTGACAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTCGACAAC
TCAGGCGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGA
AGACAGGCTGGGTCAAGTACAAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGCAATACATCAA
GGCTAACTCTAAGTTCATTGGTATCACTGAAGGCGTCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGA
ACAGGCTGGTACTACCTCAAACCAGACGGAACACTGGCAGACAGGCCAGAAATGACACCGGGCAGCCAGTCTC
CTTTCTTCTGCTGCTGCTCCTCACAGTGCTTACAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCAGG
TGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCAGCTCTACTGAGAAGAATGCTGTGAGTATG
ACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACTCAGGGACAGGATGTCACTCTGG
CCCCGGCCACGGAACCAGCTTCAGGTTAGCTGCCACCTGGGGACAGGATGTACCTCGGTCCCAGTCAACAG
GCGAGCCCTGGGCTCCACCACCCCGCCAGCCACGATGTACCTCAGCCCCGGAACAAGCCAGCCCCGGGC
TCCACCGCCCCCCCAGCCACGGTGTCACTCGGCCCCGGACACCAGGCCGCCCCCGGGCTCCACCGCGCCCGCAGCCCCAGGTGT
CAGCCCACGGTGTACCTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCACCGCGCCCGCAGCCCCAGGTGT
CACCTCGGCCCCGGACACCAGGCCGCCCCGGGCTCCACCGCCCCCCCAGCCCATGGTGTACCTCGGCCCCG
GACAAACAGGCCCCGCTTGGCGTCCACCGCCCCCTCAGTCCACAATGTACCTCGGCTCAGGCTCTGCATCAG
GCTCAGCTTCTACTCTGGTGACAACGGCACCTCTGCCAGGGCTACCACAACCCAGCCAGCAAGAGCACTCC
ATTCTCAATTCCCAGCCACCACTCTGATACTCCTACCACCTTGCCAGCCATAGCACCAGACTGATGCCAGT
AGCACTCACCATAGCACGGTACCTCCTCTACCTCCTCCAATCACAGCACTTCTCCCCAGTTGTCTACTGGG
TCTCTTTCTTTTCTGCTCTTTTTCACATTTCAAACCTCCAGTTTAATTCTCTCTGGAAGATCCCAGCACCGA
CTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGAGATTTATAACAAGGGGGTTTTCTGGGC
CTCTCCAATATTAAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCTGGCCTTCCGAGAAGGTACCATCA
ATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACGGAAGCAGCCTCTCGATATAACCTGACGATCTC
AGACGTACAGCGTGAAGTGTGTCATTTCTTCTGCCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGCATC
GCGCTGCTGGTGTGGTCTGTGTTCTGGTTGCGCTGGCCATTGTCTATCTCATTGCCTTGGCTGTCTGTCACT
GCCGCCGAAAGAACTACGGGCAGCTGGACATCTTTCCAGCCCGGATACCTACCATCCTATGAGCGAGTACCC
CACCTACCACACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTAGCCCTATGAGAAGGTTTCTGCA
GGTAATGGTGGCAGCAGCCTCTCTTACACAAACCCAGCAGTGGCAGCCACTTCTGCCAACTTGATAG

WO 03/104272

PCT/EP03/06096

45/45

FIG. 29B. ss-CPC-MUC1 protein sequence Polypeptide sequence (SEQ ID NO.52)

MGWSCIIILFLVATATGVHSQVQMAAAYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDN
SGEMATGWKKIADKWYYFNEEGAMKTGWVKYKDTWYYLDAKEGAMQYIKANSKFIGITEGVMVSNAFIQSADG
TGWYYLKPDGTLADRPENTPGTQSPFFLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSM
TSSVLSSHSPGSGSSTTQGDVTLAPATEPASGSAATWGQDVTSPVTRPALGSTTPPAHDVTSAPDNKPAPG
STAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTSAPDTRPAPGSTAPPAHGVTSAP
DNRPALASTAPPVHNVTASGSASGSASTLVHNGTSARATTPASKSTPFSIPSHHSDTPTTLASHSTKTDAS
STHHSTVPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNSSLEDPSDYYQELQRDISEMFLQIYKQGGFLG
LSNIKFRPGSVVVQLTLAFREGTINVHDTVETQFNQYKTEAASRYNLTISDVSVDVPPFFSAQSGAGVPGWGI
ALLVLVLCVLVALAIVYLIALLAVCQCRKKNYQGLDIFPARDTYHPMSEYPTYHTHGRVPPSSSTRSPYEKVSA
GNGGSSLSYTNPAVAATSANL